RADC-TR-82-270
Final Technical Report
June 1983

STEREO RECONSTRUCTION STUDY

Stanford University

Sponsored by
Defense Advanced Research Projects Agency (DOD)
ARPA Order No. 4302

Staff of the Artificial Intelligence Laboratory

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-82-270 has been reviewed and is approved for publication.

APPROVED: JOHN T. BOLAND
Project Engineer

APPROVED: JOHN N. ENTZMINGER, JR.
Technical Director
Intelligence & Reconnaissance Division

FOR THE COMMANDER: John P. Huss
JOHN P. HUSS
Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (IRRA), Griffiss AFB NY 13441. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

STEREO RECONSTRUCTION STUDY

Staff of Artificial Intelligence Laboratory
Thomas O. Binford

Contractor: Stanford University AI Laboratory
Contract Number: F30602-78-C-0083
Effective Date of Contract: 15 July 1981
Contract Expiration Date: 30 April 1982
Short Title of Work: Stereo Reconstruction Study
Program Code Number: 1L10
Period of Work Covered: Sep 81 - Apr 82

Principal Investigator: Thomas O. Binford
Phone: 415 497-2797
Project Engineer: John L. Boland
Phone: 315 330-2498

Approved for public release; distribution unlimited.

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by John L. Boland (IRRA), Griffiss Air Force Base NY 13441 under Contract F30602-78-C-0083.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER RADC-TR-82-270	2. GOVT ACCESSION NO. AD-A138208	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) STEREO RECONSTRUCTION STUDY		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report Sep 81 - Apr 82
		6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) Thomas O. Binford Staff of Artificial Intelligence Laboratory		8. CONTRACT OR GRANT NUMBER(s) F30602-78-C-0083
9. PERFORMING ORGANIZATION NAME AND ADDRESS Stanford University AI Laboratory Computer Science Dept Stanford CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62711E D30200P2
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd Arlington VA 22209		12. REPORT DATE June 1983
		13. NUMBER OF PAGES 268
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Rome Air Development Center (IRRA) Griffiss AFB NY 13441		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: John T. Boland (IRRA)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Stereo Reconstruction
Image Matching
Image Correlation
Artificial Intelligence
Mapping Systems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report surveys stereo mapping systems and related and supporting techniques. Included are these topics: Current stereo mapping systems, model based image analysis systems, image segmentation, image registration, correspondence constraints for stereopsis, hardware and architecture for computer vision, psychology and neurophysiology of vision.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

Organization of the Survey

This is a survey of stereo mapping systems and related and supporting techniques and literature. It is comprised of 8 chapters:

1. Current Stereo Mapping Systems
2. Survey of Model-Based Image Analysis Systems
3. Image Segmentation
4. Image Registration
5. Correspondence Constraints for Stereopsis
6. Hardware and Architectures for Computer Vision
7. Psychology and Neurophysiology
8. Bibliography

The general organization of the various chapters is one of presenting an overview of the area discussed, followed by a general critique of approaches and techniques, and then summaries and critiques of specific notable papers. The extent of these latter summaries and critiques varies widely in this document; some articles have insufficient content to warrant more than cursory review; for those papers considered exceptionally important, we have tried to make the detail of our commenting mirror the value of the paper; in many cases where discussion seems brief, comments will be expanded later, as the maintainance of this document will be an on-going sideline to our research.

Chapter Coverage

The following seven chapters cover a substantial part of the literature impacting on computer stereo vision research. They range over current stereo mapping and model-based image analysis systems, through techniques of edge segmentation and correspondence determination, to issues of hardware implementation of algorithms and the psychological bases of computer based visual processing.

The chapter on existing stereo mapping systems discusses the contributions of both commercial and research laboratory groups to automated stereo mapping, comments on the basic correspondence problems addressed in these works, the effect their goals have had on the techniques developed, presents a general critique of these efforts, and lays out some important points for consideration in the design of future automated stereo mapping systems.

The second chapter surveys and critiques the state of the art in model-based image analysis systems, emphasizing the progress being made toward a *general* vision system, and outlining principles felt to underlie the development of such general vision systems.

The chapter on image segmentation deals with edge detection in digital images (both local and global techniques), region growing, statistical techniques and curve segmentation, texture analysis, and global grouping operations.

Chapter four addresses the issue of image registration for both motion and binocular image sequences, and in the process discusses some relevant work in optic flow.

The fifth chapter highlights some of the constraints and parameters utilized in correspondence determination for computer vision and reviews some of the more interesting analyses in this area.

The chapter on hardware and architectures for computer vision systems draws out the distinction between the low-level relatively simple computations required for certain vision tasks, such as image segmentation, and the more complex challenges of facilitating the higher level computations for such tasks as symbolic reasoning and model matching. The structure of such algorithms leads to different demands on the architecture of the implementation, from distributed array processing to parallel processor streams; this chapter reviews a selection of the published works in this area, and provides a discussion of the relevant issues.

Chapter seven highlights some of the evocative and influential results from neurophysiological and psychological studies of human perception, categorizing them in ways that may make more obvious there impact on computer models of perception.

The last chapter is a complete bibliography for the survey. A list of cited references is included at the end of each chapter.

Contributors

The following researchers have made contributions to this survey:

Commercial Stereo Systems	Sidney Liebes
Research Stereo Systems	Harlyn Baker Sidney Liebes
Model Based Systems	Tom Binford
Edge Detection	Peter Blicher David Marimont
Texture	Ramakant Nevatia
Grouping Operations	David Lowe
Image Registration	Ken Clarkson Harlyn Baker
Correspondence Constraints	Harlyn Baker
Hardware and Architecture	Allan Miller
Psychology and Neurophysiology	Harlyn Baker Ted Selker

Chapter 1

CURRENT STEREO MAPPING SYSTEMS

This treatment of automated stereo systems is divided into two sections, one dealing with existing commercial systems and the other with research systems. The basic problem addressed in both of these is the same: to determine the distance to various locations in a scene by selecting corresponding points in a stereo pair of images and doing the simple geometric triangulation; the difficulty lies in selecting corresponding points. Commercial systems have been developed to automate terrain mapping; research systems aim for this and, further, to assign symbolic interpretations to the scene components.

1.1 Commercial Stereo Systems

The development of commercial stereo systems has been driven by requirements of terrestrial cartography. Cartography traditionally concentrated on mapping of terrain, producing elevation contour maps and digital terrain data bases. The bulk of stereophotogrammetry is accomplished by humans who perform stereo correlation or who assist interactive systems with automated stereo correlation which require extensive operator intervention. All of these systems use cross-correlation. To the extent that cross-correlation is limited, they all have similar limitations. These limitations are in mapping accuracy and applicability to various terrain types (see, for example, Ryan, Gray, *Hunt [Ryan 79] and Ryan and Hunt [Ryan 80]).*

1.1.1 – Overview

The first generation of interactive partially-automated stereo compilation systems use analog cross-correlation of small image areas (recently digital correlation) to track corresponding areas in image pairs. All of the systems to be described require distinctive texture within the area of correlation; they break track on sand, concrete, snow, or roofs (ambiguous texture or featureless area); they break track in trees (ill-defined depth). The systems break track where there are occlusions, hence no corresponding areas in two images, that is, at buildings and thin objects (poles) where the correlation area crosses surface discontinuities. Thus, the systems break track where there is no local correlation (zero signal and where two images do not correspond) or where the correlation is ambiguous (where the signal is repetitive). The systems must be started manually and corrected when they break track.

U. V. Helava invented the analytical plotter [Helava 57, Friedman 80p. 703] while working at the National Research Council of Canada. The plotter utilized computers and special servomechanisms to move the photographic plates in a manner that compensated for central projection and other factors. Construction of the first analytical plotter, the AP-1, was a collaborative effort [Friedman 80, p. 703]. Ottico Meccanica Italiana (OMI) built the optical-mechanical components, Bendix Research Laboratories had responsibility for the computer, electronic interface, and computer programs, the National Research Council of Canada consulted on the development, and the

U.S. Rome Air Development Center provided the financial support through a contract for development. First delivery was made in 1961. A succession of refinements followed [Scarano, private communication]. The AS-11A, in'roduced in 1962, exhibited 5-micron accuracy. Automated correlation was incorporated into an experimental AS-11A in 1965. Today's state-of-the-art instrument, the AS-11B-X Automated Stereomapper [Scarano 76; and private communication], incorporates laser beam epipolar scanning, address modification to implement the digital equivalent of scan shaping, and digital correlation to extract elevation data.

In 1960, Ramo Woolridge was av rded a contract to develop for the U.S. Army Engineering Geodesy, Intelligence Mapping Research and Development Agency a prototype system for the automatic production of altitude data and orthophotos from a stereo model as projected on a Kelsh plotter. This effort resulted in the production by Thompson Ramo Woolridge of the Automatic Stereo Mapping System [Bertram 63], and ultimately by Bunker-Ramo Corporation of the Universal Automatic Map Compilation Equipment (UNAMACE) [Bertram 65, Bertram 69]. These systems incorporated digital computers to position, electronically scan, and automatically estimate by correlation corresponding points in a stereo pair of diapositives. A digital phase sensitive detection technique accomplishes locking in and tracking over the 3-D model. A prototype version of this automated stereo mapper spun a Nipkow disc (the original TV scanner consisting of a disc with many small aperatures arranged in a spiral) over a rectangular window in the plane of the stereo model on the Kelsh. Computer controlled tilt of the disc accomplished relative stereo window warping to accommodate terrain tilt. In later versions the Nipkow disk was abandoned, partly because it contributed an unacceptable degree of vignetting. Thereafter, window warping was not a feature of the system.

During this same period, Gestalt International Limited produced the GPM-1 and GPM-2 Gestalt Photomapping Systems [Kelly 77, Allam 78]. The GPM-2 computer-controlled, autocorrelating, analytical photomapper represents a high level in the state-of-the-art in automated mapping. The instrument is designed to operate upon hard copy input, although it could presumably be redesigned to operate from a digital data base. Scanning is accomplished by flying spot scanners, reducing optics that focus the spots onto the input diapositives mounted on transporter stages, and photomultiplier recorders. The system would appear to be well suited to typical natural terrain analysis, but encounters the same problems experienced by all related area cross-correlation systems when confronted with steep and overhanging relief, such as that frequently encountered in dealing with common cultural artifacts.

Steve Mildenberger has kindly called to our attention that during the period 1972-76 information was collected by the International Society of Photogrammetry on the quality of products generated by commercially available orthophoto systems [Blachut 76]. Galileo, Gestalt, Kelsh, Ortho, Wild, and Zeiss instruments were involved in the test. As the test was rather of orthophoto production capability rather than automatic correlation, Unamace and AS-11 equipment was not inclued. Participating manufacturers were provided with a stereo pair of 1:10,000 contact diapositives. A photograph of the test area selected for the experiment shows rolling farmland, a portion of which is covered by dense forest, some isolated trees, what appears to be a dammed section of a river, a rural road structure, and a few small buildings. It was noted [Ibid. p. 4] that whereas "many open hills in the test area are particularly suited for evaluating the accuracy of the planimetric and height information provided by the orthophoto technique ... [the] buildings and isolated groups of trees in the area offer a possibility of evaluating the influence of these 'elevated' features n the geometric quality of orthophotos and contour lines, particularly when using automatic image correlation."

The manufacturers were asked to produce the orthophotos and the height data with their own equipment. Work was evaluated by the National Research Council of Canada. The only systems incorporating automated image correlation were the Gestalt International Limited GPM-1 and the

GPM-2 (the latter being an improved version of the former.) As to be expected, the automatic correlation of the Gestalt systems yielded deteriorated elevation and rectification data in the vicinity of elevated features such as trees and buildings. It was concluded that point elevations within 10 meters ground scale (1 mm on the original photos), should not be used in practical projects, at this scale. Omitting errors generated within 1C meters of buildings and trees, the accuracy of the GPM-2 was judged comparable with that of non-automated correlation equipment. Its speed of performance in preparation of the control manuscript and the relative and absolute control procedures was three times faster than that for conventional equipment (37 minutes vs. an average of 1.5 hours for the nine or so pieces of equipment involved in the testing). The GPM-2 is suggested to run approximately 50 percent faster in the production of orthophotos, along with recorded and plotted terrain heights, than do non-automated correlation systems. It is remarked at the end of the report that stereo orthophotos submitted by Gestalt International Limited arrived too late to be evaluated and that analysis of them would be carried out in a separate study. We have not seen the subsequent study, but presume that the results of further analysis would not appreciably influence the general conclusions regarding performance of the automatic correlation capability reported here.

1.2 Research Stereo Systems

The objective of research stereo systems is the development of accurate, robust, totally autonomous three-dimensional mapping for input to object or terrain modelling systems. The principal distinction between these and commercial systems is in their exploitation of both geometric and photometric constraints for their analyses. Such systems run on large digital computers with considerable amounts of memory. The large memory allows access to image data at rates limited by electronics rather than mechanics. Some use the semantics of two-dimensional structural entities in the scene to guide the search for correspondences and restrict the set of depth determinations — requiring global consistency. Being experimental, they tend to work on much smaller images than the commercial systems described earlier. None as yet runs in real time, although coming generations of parallel mechanisms are expected to provide the computational power to enable this.

1.2.1 – Overview

Given a pair of corresponding points in two views of a scene, when the relative position and orientations of the two imaging sites is known, the depth to the viewed spot can be obtained through a simple triangulation. The real problem in stereo analysis is in determining those pairs of corresponding points. Two main categories of approaches exist in this stereo correspondence work, and the difference centers on their use of ***feature-based*** or ***area-based*** correspondences. The ***feature-based*** systems to be described are those of [Arnold 78], [Grimson 80], [Baker 81b], and [Arnold 82], and the ***area-based*** systems are those of [Gimel'farb 72], [Levine 73], [Mori 73], [Hannah 74], [Panton 78], [Gennery 80], and [Moravec 80].

The distinction between 'feature' and 'area' correspondence here can be more a matter of degree than type. ***Feature-based*** analysis involves the transformation of the sensed data from a discrete two-dimensional intensity array to a more symbolic form as significant intensity contours, or 'edges' - features. It is the properties of these features which then provide the metric for the correspondence. 'Feature' is a fairly general term, but its use here may be equated with 'edge'.

There are many fewer 'edges' than image elements in a view of a scene, so this transformation, generally, reduces the computational cost of determining correspondences. A drawback is that not every point in an image is a 'feature', so the result of a solely feature-oriented correlation will not be the dense depth map one may want.

In *area-based* analysis two-dimensional windowing operators measure the similarity in intensity pattern between local areas, or windows, in the two images. Cross-correlation is used to determine matches between windows in one image with windows in the other. Normalized cross-correlation has the ability to compensate for contrast and brightness differences across images. If lighting and sensor/processing conditions are known, this flexibility in the algorithm may not be required. In this case other correlation forms (such as normalized RMS, or absolute difference) may be used. Area cross-correlation is often not applied to every pixel in the image arrays, but selectively for those whose local variance is high. Variance measures have been used as filters to limit possible correspondences, with correlation being used to select the best from among the candidates. These variations may qualify such approaches as 'feature-based', although they will not be considered so here. Perhaps a better way of categorizing these systems is as *feature-driven area-based* ([Moravec 80] uses an 'interest operator' to select worthy points in a reference image, [Henderson 79a] preprocesses the data to find edges which are then used to bound the area-based search, [Levine 73] limits initial correlation to areas having local maximal variance, and [Gennery 80] uses a variance based F test to filter out areas of minimal information, and therefore minimal interest).

1.2.2 – Area-based processing

Area-based correspondence has been applied quite successfully to the stereo analysis of rolling terrain, but it degrades when the scene is not smoothly varying and continuous. In images of such domains many windows to be matched will have no correspondences in the other image (for example, those windows lying on surfaces which are occluded from the other imaging position). The chief difficulty with the area-based approach is in properly matching window shapes and sizes for conjugate image areas ... taking into account both variation in terrain slope and discontinuities at surface boundaries.

Large correlation window sizes are required in attaining statistical significance in the sampling, yet the characteristics measured over the windows become less and less representative of the observed local surface as this window size increases. Discontinuities in the surface can cause a positioned window in one image to be sampling local intensity values from more than one intensity surface in the other image, and a correct cross-correlation would only be possible if the window could be partitioned and matched with (possibly several) windows of various size and shape in the other image. Such adaptation requires more flexibility than area-based correspondence has thus far been shown to provide. Abrupt discontinuities in topographic structure and an abundance of occlusions characterize urban or cultural areas. It is at precisely these points of depth discontinuity that we want to obtain accurate surface position measurements. This would suggest that current area-based processing is inappropriate for domains with occlusions and abrupt depth discontinities.

Some consideration of this window shaping problem has been attempted in area-based work. Levine and O'Handley ([Levine 73]) and Mori, Kidode and Asada ([Mori 73]) vary their correlation window sizes with the local intensity variance. They presume that high variance implies high local texture and thus suggests the need for smaller correlation windows, while low variance suggests surface uniformity and the need for larger sample sizes and larger correlation windows. [Panton 78] uses trapezoidal window shapes in the search image, as determined by previous and predicted correspondence results, to match the rectangular windows of the reference image. [Gennery 80] included a partial solution to this problem for a specific camera geometry when looking at windows presumed to lie in the ground plane. Mori, Kidode and Asada implemented an iterative technique that would compensate for terrain variations by successive refinements to image registration estimates. Both [Levine 73] and [Hannah 74] included in their algorithms techniques for identifying scene occlusions and areas of image non-overlap, but these were entered as cases of exception handling, and it is doubtful that they were adequate as models of occlusion.

A related problem with area-based correspondence is that increasing window size improves statistical significance but generally results in poorer 3-space positioning accuracy for the correspondence. Feature-based analysis obtains more precise positioning (for its edges) in the individual images, and it can attain correspondingly higher accuracy for its correspondences in 3-space ([Arnold 78] indicates that edge-based techniques offer a factor of 10 improvement in accuracy over area-based correlation methods).

Area-based correspondence systems also tend to be prediction driven, in that they process an image serially and at each step use the context of previously matched neighbouring points to limit the search for the present correspondence. No backtracking facility is provided with this technique, and only [Gennery 80] includes a capacity for correcting locally determined miscorrespondences. With little ability to either correct or detect errors, such a prediction approach can lead to rapid degeneration once errors begin to occur.

A final, and important distinction between area-based and feature-based processing lies in the basic philosophy of their analyses. The underlying assumption of area-based correspondence is that it is the photometric properties of a scene that are invariant to imaging position, and the correlating of these properties will be sufficient to allow the proper correspondences to be determined. But it is not the measurable photometric properties that are invariant to viewpoint positioning. In the degenerate, although common enough case, a surface of a certain intensity seen unobscured from one viewpoint will not even be visible from another slightly different viewpoint. All that can be said to be *truly* invariant to viewpoint positioning is the three-dimensional structure of the scene itself. A better metric for the correlation would be one which deals in some way with that scene three-dimensional structure.

1.2.3 – Feature-based processing

Feature-based analysis, in the form of edge analysis, comes closer to dealing with this scene structure invariance. It works generally with the premise that a local measure on the intensity function is representative of physical change in the underlying scene. The local measure on the intensity function could be, for example, a maximum in intensity gradient — peak in the first difference of intensity, zero-crossing in the second difference. Physical change in the scene could be a break in depth continuity and accompanying projected surface reflectance or luminance change, or a change in surface intensity from a surface detail without topographic break. The point to notice is that feature-based analysis uses the semantics of intensity variation in its attempt to extract measures of the physical change in the underlying structure of the projected views, and uses these two-dimensional observations to infer the three-dimensionality of the scene. The validity of this intensity edge tracking in a stereopsis system is apparent:

- a discontinuity in surface orientation will, in general, give rise to a variation in incident reflection, which will appear to an imaging source as a change in brightness – tracking the intensity edge across the two views will track the surface discontinuity;

- an illumination discontinuity .(shadow edge), although not likely corresponding to a surface discontinuity (the shadow will lie *on* the surface), will be visible as a brightness discontinuity – tracking the shadow edge across the two views will provide depth information about the shadow-bearing surface;

- surface marking or pigment variation will similarly provide depth information along the bearing surface.

Probably the most widely known edge-based stereo scheme to date is that of Marr and Poggio ([Marr 77]), as implemented in a computer program by Grimson (see the following summary [Grimson 80]). The algorithm has been fairly well tested on a reasonably wide variety of images (random dot stereograms, natural terrain, urban scenes), and is at present being implemented in hardware [Nishihara 81]. The stereo processing system of Henderson, Miller and Grosch of the Control Data Corporation research group ([Henderson 79a, Henderson 79b]), called the Automatic Planar Surface System, uses edges to guide it's area-based matching. They address their work specifically toward the problem of constructing planar models of rectilinear cultural scenes from stereo pairs of aerial imagery. [Arnold 78] developed an edge-based stereo correspondence system that used local edge properties to select edge match possibilities, and a weighted iteration process to resolve match conflicts. Baker's approach ([Baker 81b]) incorporates both edge and intensity based correspondences, although unlike the CDC algorithm it actually correlates the edges. It uses the relatively sparse results of its edge analysis as a template for a fuller pixel intensity matching process. An extension of the CDC work ([Panton 81]) has lead to a stereo correlation system that uses both the *local* edge information (as the above systems) and *extended* edge information in its stereo matching. Recent work by Arnold ([Arnold 82]) has resulted in a line-by-line edge-based correspondence scheme that uses extended edge connectivity to disambiguate match rivalries. The line-by-line analysis retains the optimal as well as several near-optimal alternate sets of correspondences.

1.3 Stereo System Summary

1.3.1 – Autonomous processing

A stereo system to operate for mapping, reconnaissance, or inspection in some domain must be able to initialize itself and run without the need of operator intervention. Of the recent systems summarized, only Gennery's and Baker's are designed to run entirely autonomously. Panton's appears to require manual initialization, as does certainly the Henderson, Miller, Grosch system ([Henderson 79a, Henderson 79b]) and, to a lesser extent, the Grimson system ([Grimson 80]). These may also require manual intervention during the processing — the Henderson system when there are vertical breaks in scene continuity, the Grimson system when the disparity differences exceed the size of the largest convolution operator, and the Panton system when the terrain approaches discontinuity and the correlator begins to diverge locally from the correct matchings.

1.3.2 – Domain restrictions

An understanding of its domain of intended use and an analysis of its performance capabilities will give us insights into a stereo system's overall range of application, and thus its utility. In general, the performance of the area-based correspondence schemes will degrade rapidly when confronted with scenes of discontinuous structure, and this makes them inappropriate for the analysis of cultural sites. The CDC techniques of [Henderson 79a, Henderson 79b] and [Panton 81] exclude the processing of rolling, curved, or even non-rectilinear structures — predisposed to the analysis of building tops, they are inappropriate for most everything else. None of the systems described can work well where details in the background have reversed positioning with respect to occluding surfaces lying before them (consider a finger at arms' length and the background beyond) — this is referred to as the *edge reversal* problem. The Grimson work is the only one which does not explicitly exclude such positional reversals between the two imaging planes (although it probably does so implicitly in the working of its region disparity consensus). Excluding edge reversals is such a convenient expedient when working with epipolar geometries that it has been widely accepted for the correspondence processings. That it is a restriction becomes obvious when it is noticed that it prohibits the simultaneous fusion of a thin object (like a pole) and its background — relative to the pole, what is left-right in one image will be right-left in the other. This artifact of the processing may be excused to some extent in that it is also observed in human stereopsis, but there is no necessity to build limitations of the human system into a machine system (in their study of the limitations of binocular fusion, Burt and Julesz ([Burt 80]) comment on the impossibility of fusion of positionally reversed points).

Looking at the range of examples presented in the published results from these stereo systems also provides insight to their applicability. [Panton 78] has demonstrated a single rolling terrain stereo pair analysis, as has [Gennery 80], although Gennery's scene contains some rather large rocks and the scene slopes off to a (not seen) horizon. Levine and O'Handley show the processing of two rock strewn scenes, similar to that of Gennery. The views in these area-based systems are, as expected, of terrain, and depth discontinuities are either not severe or ignored. Baker shows the results of his processing on two pairs of images, one cultural (and synthetic), and one of rolling natural terrain. Grimson has applied his algorithm to considerably more scenes ... many random dot stereograms, and several real image pairs. Arnold's system has been applied to the analysis of a single hand-extracted edge description of a stereo pair ([Arnold 82]).

1.3.3 – Global consistency and monocular cues

The human perceptual system has the advantage that it can call upon higher processes to comment on the consistency of its visual observations. Only rarely is our binocular sight confused by ambiguities, and then this can usually be removed with a tilt of the head or slight motion to the side for a different perspective and more information (an observation which lead Moravec [Moravec 80] to his development of *slider stereo*). An interpretation mechanism is at work with which our stereo systems at present have little to compare. Some researchers have decided that local averaging of depth measurements provides a reasonable approximation to error correcting, hoping to diminish the impact of gross errors through the abundance of good correspondences (for example [Levine 73] and [Grimson 80]). A superior approach is to work within a set of plausible assumptions on the nature of the viewed world, and use the implications of these assumptions to chose among ambiguous interpretations. A common assumption is, for example, that the world is smooth and continuous most everywhere, and can be expected to be discontinuous only at those places where the viewed luminance is undergoing abrupt change - that is, to presume that significant changes in scene brightness may separate areas of different depths, and where there are not such brightness changes, the surface is likely to be smooth and continuous.

The way such knowledge enters the analysis varies. In some work, the continuity assumption is used in prediction. Levine, Panton, and Gennery, in their area-based systems, use the context of neighboring points to limit the search for point correspondents, presuming that points neighboring in two dimensions should be neighboring in three dimensions. But this has problems of consistency - the results would change were the analysis to be done right to left rather than left to right, and decisions are made locally, in a set direction, never to be revised. The MIT Grimson work makes good use of inference on the continuity of surfaces and the lack of edge signal in its interpolated surface fitting (see the summary). However its use of context in its local edge correlation is only cursory, in that matching at a lower resolution (lower spatial frequency) is a prerequisite for matching at a level of finer detail (higher spatial frequency). A global metric is used in consistency checking over regions — requiring 70% of the disparities to be in agreement (one standard deviation, presumably), but this has been implemented without adequate analysis (see [Grimson 80] page 213, where it appears to produce a highly quantized, planar effect). [Schumer 79] discusses a possible mechanism in the human system for this spatial averaging of disparities. The Baker algorithm uses a similar global continuity assumption arising from two-dimensional connectivity to resolve disparity ambiguities along connected contours — the assumption being that projected connectivity in two-dimensions is a good indicator of connectivity in three-dimensions, and abrupt changes in disparity along a path connected in 2-D suggests a correspondence error.

1.3.4 – Identifying depth discontinuities

An issue related to the use of global continuity assumptions is the identification of depth discontinuities in the scene - those places where the viewed surface is not smooth and continuous. This capacity has not been reliably incorporated into area-based analyses, where poor matches arising from occlusions or extreme perspective effects merely return a low correlation value, indistinguishable from other causes of poor matches. In cases of occlusions, the intensity values in a window about the depth discontinuity in the two views would have little likelihood of corresponding, and the correlation coefficient as a measure of similarity is inappropriate here. Edge-based analyses operate with the artifacts of (among other things) depth discontinuities, and the inference is available here for distinguishing occlusions and abrupt changes in depth. In [Baker 81b], edges found by a zero-crossing operator are split into their left and right components, and these 'half-edges' are matched across images. Contrast variations (even reversals) needn't affect the correlation. Surface matching (based primarily on intensities) is performed within intervals defined by pairs of such corresponding half-edges - occluded regions will not have pairs of corresponding half-edges (Baker points out in [Baker 81b] that the algorithm does not use two-dimensional global information in its pixel intensity correlations - smoothness between space contours is only enforced along the baseline axis - and this should be altered in a better solution).

In an early interesting variant on area-based correspondence, [Levine 1973] used a technique of mixing area-based correspondence with feature-based guidance. Image points having locally maximal directional intensity variance were selected and correlated first. These they referred to as 'tie-points'. The disparity values for these tie-points then limited the search range for neighboring image points, and certain heuristics were introduced to infer and compensate for occluded regions. The tie-points also provide a disparity context for the matching of nearby image elements. [Mori 1973] alludes to a similar technique, saying that they process first those parts of the images having high intensity contrast, and use the disparity determined there, with an assumption of scene continuity, to constrain the disparity of neighboring pixels.

1.3.5 – Parallelism possible

A stereo system to be used for tasks of navigation or process control must be judged on its ability to provide depth measurements at rates approaching real-time. The enormous amount of computation inherent in the analyses makes it unlikely that a scheme with intrinsic ordered dependence in its processing will be able to provide adequate speed. The potential for parallelism in the algorithm is an important consideration. Neither the Gennery nor the Panton approaches could take full advantage of the high parallelism possible in the computation since they process from left to right in columns across the match image, relying upon previous correspondences to constrain the search for matches. The Henderson and Levine approaches are similarly limited, in that they process by lines from image bottom to top, with each line progression passing up the results of the preceding line analyses to constrain the search. The Grimson and Baker algorithms are both amenable to parallel implementation.

1.3.6 – A Composite Solution

The principal problem with the area-based cross-correlation approach lies in the difficulty of shaping image patches for the matching across images. Projective distortions and occluding surfaces exacerbate the problem of chosing frames for the comparison of conjugate areas. This window shaping difficulty makes area correspondence fail in areas of varying relief and it fails wherever there is occlusion (i.e. at depth discontinuities).

Edge-based correspondence encounters difficulty when edge density or lack of local distinguishing edge characteristics make for ambiguous matching. More global characteristics (either in the projective images, or over time) could resolve the correspondence. The globality here could come from connectivity over extended edges (as used in [Baker 1981b], [Panton 1981], and [Arnold 1982]) or contextual cues as discussed in [Binford 1981] — these utilize monocular cues to stereopsis. It could also come from stereo observation over a series of views — either spatially, as in [Moravec 1980] using 9 views of the scene, or temporally, with a focus on ambiguous areas and a feedback loop to adjudicate the rivalry.

The matching of extended edges will also not be without ambiguity. Occlusions will disrupt projective continuity, and the matching of these extended edges need not be one-to-one. The evidence of local edge correspondences could help in the resolution of these conflicts.

What should be seen is that there will be ambiguities – they can never be guaranteed to be resolved within a certain *spatial* frame, or within a certain *temporal* frame.

- *local edge* matching and *extended edge* matching can help each other in a pair of views;
- observations qver *time* or over *multiple views* can assist both;
- access to a *model memory* can be used to guide, or disambiguate surface interpretation;
- *edge* correspondence provides the surface boundary and surface orientation context that eludes and invalidates area-based cross-correlation approaches;
- *area-based* correspondence, where it is applicable, away from surface boundaries, specifies the shape of the surface ... edge correspondence only specifies the position of surface boundaries or interior detail.

A stereo mapping system expected to process images of both cultural and rural scenes with precision and reliability will need the benefits of each of these correspondence approaches:

- edges (binocular cues),
- extended edges and edge contexts (monocular cues),
- integration over time (resolution of ambiguity through time),
- integration over multiple images (resolution of ambiguity through redundancy),
- area-based cross-correlation (fuller surface description).

1.4 Commercial System Summaries

1.4.1 – Bunker-Ramo UNAMACE

"Automatic Map Compilation," Bertram, S., Photogrammetric Engineering, p. 184, January 1963.

A description is presented of a prototype model of an optical-electronic automatic stereomapping system that can be added to conventional plotters to produce altitude information and orthophotos automatically from pairs of aerial photographs. A pair of light sources is projected through diapositive stereo pairs so as to cross in image space at a 0.3"x0.1" window mask and thence to a pair of photomultipliers. A Nipkow disk, the original TV scanner, spins over the window, effectively scanning lines from the window. The window-disk subsystem is controlled to move as a unit through image space along the surface of the stereo model. The scanning disk tilts to conform roughly with the tilt of the terrain. Control of the window to follow the surface of the model is accomplished as follows. The output of each of the photomultipliers passes through a delay line. The delayed signal from multiplier 1 is correlated with the undelayed signal from correlator 2, and vice versa. The outputs of the pair of correlators is input to a differencing network. When the window is locked onto the surface of the stereo model, the output of the differencing network is zero. If it is off the surface of the model, the sign of the network output indicates the direction to move to return to the surface. The device amounts to a phase sensitive detector.

It is reported that "rugged terrain could be handled quite readily if a very small signal were used". Besides the difficulties to be expected over water, cliffs, and relative obscurations, the system exhibited limitations associated with the direct projection system and the Nipkow disc. Vignetting effects limited the signal in the corners of the model and at large angles of model tilt.

"The Automated Map Compilation System," Bertram, S., Photogrammetric Engineering, p. 678, July 1963.

An experimental automatic map compilation system is described. Scanning of the stereo diapositives is accomplished with flying spot scanners and movable lenses. An operator is able to monitor system performance through an electronic stereo viewer involving a twin TV system. The instrument scans in a profiling mode, with a film carriage moving in a direction nearly perpendicular to the flight line, and sampling in 0.01" steps. Profile lines are spaced 0.02" apart. The operator can intervene if the system is observed to drift off the model surface, or if alerted by a "lack of correlation light". As the diapositive carriage moves smoothly during the scan, the electronics incorporates "stop motion" circuitry analogous to that of an image motion compensation system of a camera. The sampling window is 0.050"x0.050". Whereas, the system incorporates a square scan, plans are offered for converting to a parallelogram, of a shape to be dictated by the slope of the local terrain. The sensitivity of the instrument with "good imagery" is reported to be comparable with that of a human operator. Steep cliffs, differentially hidden areas, and areas of low detail cannot, of course, be automatically processed.

"The Universal Automated Map Compilation Equipment," Bertram, S., vol. 15, part 4, International Archives of Photogrammetry, 1965; Photogrammetric Engineering, p. 244, 1965.

This paper was written at the time that the Universal Automated Map Compilation Equipment (UNAMACE) was nearing completion at Bunker-Ramo. The UNAMACE was based on the principles demonstrated in the Automated Map Compilation System, described in the earlier papers by Bertram. It is suggested as advisable, in order to address ambiguous correlation peaks that can arise from repeated structures such as orchards, to use "dual correlators, with a high acuity channel taking advantage of the available resolution in the photographs, [and a] low-acuity channel ... operating on low-resolution imagery in the photographs". The intent is that "the high-acuity channel provide tight tracking most of the time, with the low-acuity channel helping over difficult areas (such as large altitude changes)".

Correlator circuit bandwidth is cut off at the low frequency end to suppress general highlights, and the high end near the intended frequency resolution of the correlation. The UNAMACE incorporates a high-acuity and a low-acuity height-error sensor, the outputs of which are "appropriately summed" to provide a net height-error signal.

Transformation from object space to film space accounts for camera model, lens and film distortion, earth curvature and atmospheric refraction. Each stereo diapositive is mounted on a separate scanning table with its own flying spot scanner. Scanning operations are controlled by a Bunker-Ramo Model 133 computer. The system could process highly convergent photography, and distortions of panoramic cameras. The tables could accommodate 9"x18" glass plates, and translate up to 2 inches per second at 4-micron rms accuracy. Correlation tracking incorporates extrapolation from previously processed points. The high-speed scan direction (X) is parallel to the line joining the two camera stations. A TV-like coverage is input to the correlator at each sample location, with the relatively slower scan in the Y-direction. Correlators processing the Y-component of the signal are used to remove Y-parallax errors from the data.

A stereo-viewer on the control console enables monitoring of the automatic compilation. The viewer incorporates two 4 inch, square-faced CRTs. Stereo viewing is accomplished via crossed Polaroids and image merging through a beam splitter. A stereo floating mark is generated electronically.

Altitude determinations are made at a rate of 100 points per second, representing a three-fold improvement over the earlier system. At the time of Bertram's writing of this paper, the effectiveness of the UNAMACE operational cycle remained to be verified. It was anticipated that significant improvements would result from programming modifications that would permit the density of measurements to adapt to the terrain. It was estimated that the UNAMACE would compile 100% lap diapositives in about 3 hours using 250-micron spacing along the profile and 500-micron spacing between profiles.

The UNAMACE produces orthophoto products as well as elevation contour information. It was recognized the it would be desirable to generate contour lines rather than the rotary sequence of three gray tones of the system described.

"The UNAMACE and the Automatic Photomapper," Bertram, S., Photogrammetric Engineering, vol. 35, no. 6, p. 569-576 June 1969.

The UNAMACE has been found to operate at up to 80 points per second. Diapositive table control is repeatable to 2-micron, with an absolute accuracy of 4-micron.

A small ruggedized van-mounted version of the UNAMACE called the Automatic Photomapper has been developed. It is restricted to 9"x9" photographs.

1.4.2 – Bendix AS-11B-X

"A Digital Elevation Data Collection System," Scarano, Frank A., Photogrammetric Engineering and Remote Sensing, vol. 42, no. 4, p. 489, April 1976.

The AS-11B-X automated stereomapper experimental system was developed by Bendix Corporation for the Rome Air Development Center specifically to generate elevation data in digital form. It employs laser scanning along epipolar lines, analog-to-digital conversion of photo data, digital storage, and digital correlation. The elevation points are interpolated off-line to a fixed grid. The instrument performs single-pass scanning, interim random access digital storage of density values, and subsequent recall for correlation calculations. Correlation window shaping is accomplished by address modification of digitally stored imagery data. In the case of perfectly registered images, the utilization of the epipolar scan pattern anables correlation matching to be accomplished along corresponding epipolar lines. It is reported that the variance of residual parallax in the direction transverse to the epipolar lines is always small, implying that the effect of y-parallax variation can be neglected.

Image scanning is accomplished by optical-mechanical straight-line deflection of the laser beams over the images. It is thus implied that scanning of nonlinear epipolar lines is accomplished by linear approximation. The image density is sampled at 1280 points spaced at 20 micron intervals over a 25.6 mm deflection of the laser beam. The sampled points are partitioned into 58 segments each of which is associated with a separately developed profile running in a direction transverse to the laser sweep. Successive sweeps of the laser beams are displaced 50 microns apart by a mechanically stepping the film stages.

Correlation patch shaping and digital correlation are accomplished by special-purpose digital logic. Digitized imagery values are discarded as soon as each local area is correlated. It is stated that normalized correlation is used both for evaluation of system performance and determination of system stategies, but that image-correspondence determination is based on faster local covariance maximization. Regions for which the correlation threshold falls below an established threshold are filled in by a human operator.

1.4.3 – GESTALT GPM Photomapping Systems

"The Gestalt Photomapping System," Kelly, R. E., P. R. H. McConnell, and S. J. Mildenberger, J. of Photogrammetric Engineering and Remote Sensing, vol. 43, p. 1407, 1977.

The Gestalt Photomapping System is a computer-controlled, autocorrelating, analytical photomapper, combined with an off-line plotting unit. The photomapper is composed of a pair of scanners, an automatic image correlator, a control computer, an operator's console, and one or two printers. Each scanning unit is comprised of a flying spot scanner, reducing optics that focus the spot onto the input diapositive mounted on a transporter stage, and a photomultiplier recorder. The input diapositives are processed in 0.72 sq. cm. patches. Auto-correlation of an array of 2444 points (47 x 52 matrix) spanning each patch is accomplished by analyzing and digitizing the photomultiplier output, measuring parallax for each point, storing and smoothing the parallax data, calculating and implementing changes in the video scanning pattern to minimize parallax values, reshaping the scanning patterns accordingly, and then repeating the entire processing sequence until the parallax values of the individual points stabilize. This process, referred to as complete differential rectification, occurs over the square centimeter patch instead of at a floating mark, and does so in less than a second. The system produces orthophoto hard copy products, as well as magnetic tape output. Contour or profile lines may be optionally overlayed on the orthophotos. The system contour sheets differ from conventional map maunuscripts in the important respect that they contain only topographic information and not planimetric details, such as roads, buildings, et cetera.

The Gestalt Photomapping System represents a high level in the state-of-the-art in automated mapping equipment. The instrument is designed to operate upon hard copy input, though could presumably be redesigned to operate from a digital data base. The system would appear to be well suited to typical natural terrain analysis, but must be presumed to encounter the same problems experienced by all related area cross-correlation systems when confronted with steep and overhanging relief, such as that frequently encountered in dealing with common cultural artifacts.

"DTM Application of Topographic Mapping," Allam, M. M., Photogrammetric Engineering and Remote Sensing, vol. 44, no. 12, p. 1513-1520, Dec. 1978.

A description is presented of an application of the Gestalt Photomapper GPM-2/3 to DTM (Digital Terrain Mapping) by the Topographical Survey Division, Canadian Department of Energy, Mines and Resources. The GPM-2/3 is described as a highly automated stereophotogrammetry system designed to produce DTMs, orthophotos, and photographic contours by using electronic image correlation to measure parallaxes. The instrument incorporates flying spot scanners and a digital correlator module. The correlator iterates over all 2444 points in a 182-micron patch 50 times per second. The height determinations are used to differentially transform each scanner's raster until the stereo correspondence processing stabilizes over the patch.

1.4.4 – Stanford Viking Mapper

(a computer-based, non-commercial stereo photomapping system)

"Viking 1975 Mars Lander Interactive Computerized Video Stereophotogrammetry," Liebes, Jr., S. and A. A. Schwartz, Journal of Geophysics Research, 82, no. 28, p. 4421, Sept. 30, 1977.

A computer aided stereophotogrammetry system was developed by Liebes and Schwartz in support of the Mars Viking 1975 lander camera system. Whereas the stereo correlation was performed by a human operator, the system exhibited a versatile stereophotogrammetry capability. As the system was developed outside the mainstream of the photogrammetric community, it is not widely known. Digital stereo images radioed back to Earth from Mars were stereoscopically viewed on a pair of high resolution video monitors of a Stereo Station. Visual stereo model formation was complicated by the combination of spherical coordinate scannning geometry of the lander cameras, differential stereo illumination conditions, image acquisition resolutions that could differ by a factor of three, and near field, high-convergence stereophotogrammetry requirements. Stereo model formation was facilitated by independent orthogonal image size, brightness and contrast controls, and image rotation via motorized rotation of the monitor deflecting yokes. A graphics 3-D point mark was overlaid into the imagery prior to routing of the visual data to the video monitors. Thus, analogue image warping at the monitors had no effect on the relationship of the 3-D mark to the imagery. The 3-D mark could be moved through the perceived Martian relief by hand input control. Individual points could be ranged, and profiles of arbitrary character could be generated by selection of the family of surfaces of profile constraint. Parameters relating hand input control position to 3-space mark location could be varied to suit operational conditions. The system had real-time editing facilities, could optionally display the profile graphics as stereo image overlays, stereoscopically reproject the profiles from arbitrary viewing directions, and output the topographic data in a variety of formats and scales.

1.4.5 – Published Reviews

[Friedman 80, p. 699-722] presents a review of automation of the photogrammetric process. The review covers mensuration, rectification and orthorectification, elevation determination, planimetry delineation, and integrated systems, as well as a lengthy section on the history, principles, and current designs, and also a section devoted to methods of automatic image correlation. He reports [Ibid., 701] that a thorough description of on-line automated elevation determination systems is presented in [Dowman 77]. We have not, however, been able to secure a copy of this report. Friedman also reports that the entire November 1978 issue of Photogrammetric Engineering and Remote Sensing is devoted to automated analytical stereoplotters; however, we find this reference to be erroneous. Perhaps the intended pointer was to the December issue, which contains discussion of the research work of CDC [Panton 78], and applications of the Gestalt Photomapper [Allam 78], and a Galileo Santoni 2C Orthophoto System manufactured by the Galileo Corporation of America. No literature references are given to the Galileo system. Automatic orthorectification is reported to be achieved by the Wild B-8 Stereomat, and the Jena Topomat, as well as by the UNAMACE and the Gestalt GPM-2. Friedman [Friedman 80, p. 701, 719] also reports that Horbrough [Hobrough 78] has proposed another digital on-line correlation concept, called RASTER. The particular novelty appears to be in the use of a solid state linear array scanner of 1728 elements, rotated by servomotors to lie substantially along epipolar lines. Friedman remarks [Friedman 80, p. 701] that the future of elevation determination and rectification may well lie with off-line digital image processing. Off-line processing is exemplified by the research work of Panton (et. al.) at CDC, and, though not noted in Friedman's review, by the work of Arnold, Baker, and Binford at Stanford University, Grimson and Marr at MIT, and others, as described elsewhere in this survey.

U. V. Helava [Friedman 80, p. 703] is credited, while working at the National Research Council of Canada, with the invention of the analytical plotter [Helava 57], used for non-automated, but computer assisted correspondence computations. The distinction relative to what he refers to as the "simulation principle" is that in simulation systems real physical projection is employed. Helava cited as disadvantages of the simulation approach the inflexible, bulky, extremely high precision instruments required. In the analytical plotter, computers and special servo-mechanisms move the photographic plates in a manner that compensates for central projection, and other factors. The first analytical plotter AP-1 was built by the NRC. The bulk of the analytical plotters on the market today do not incorporate automated correlation.

Friedman, in addition to discussing the principles of the analytical plotter [Friedman 80, p. 704], describes eight current designs [Ibid., 709] that were presented at the Analytical Stereoplotter Workshop and Symposium in Reston, VA, in April of 1980. None of these systems employ automated correlation. These instruments and their manufacturers are as follows:

Analytical Photogrammetric Processing System IV (APPS-IV)
Manufacturer: Autometric, Inc, Falls Church, VA.

Analytical Stereoplotter Model AP/C4
Manufacturer: Ottico Meccanica Italiana S.p.A. (OMI)of Rome, Italy.

Autoplot
Manufacturer: Systemhouse, Ltd., Ottawa, Canada.

C100 Planicomp Analytical Plotter
Manufacturer: Zeiss, Oberkochen, Germany.

DSC-3/80 Analytical Stereoplotter
    Manufacturer: Keuffel & Esser Co., San Antonio, Texas.

Galileo D.S. Digital Stereocartograph
    Manufacturer: Officine Galileo S.p.A. of Florence, Italy.

Traster Analytical Stereoplotter
    Manufacturer: Matra Optique, Rueil-Malmaison, France.

US-2 Analytical Stereoplotter
    Manufacturer: Helava Associates, Inc., Southfield, MI.

Additional analytical stereoplotters that are not well publicized, because of their highly specialized character, but which are stated to "represent large investments, produce tremendous amounts of photogrammetric data and are examples of the latest technology" are also cited [Ibid., 715]:

AS11/A1
    Ottico Meccanica Italiana (OMI), Rome, Italy
    Control computer manufacturer: Modular Computer Corp.
    Interface manufacturer: O.M.I. Corporation of America

AS11/A1(NOS)
    Interface: PDP 11/45
    Manufacturer: OMI

TA3/P1 & TA3/P11
    Control computers: PDP 11/60
    Interfaces: Bendix Research Laboratories
    Principal users: U.S.D.M.A., U.S.G.S.
    Manufacturer: OMI

Another analytical plotter, referred to only in passing [Ibid., p. 701], is the Anaplot. The entire June 1979 issue (pp. 89-192) of the Canadian Surveyor is devoted to this instrument, which Blachut (p. 89) refers to as "in many ways, ... the most advanced system of its kind in operation." The instrument does not incorporate automated stereo correlation. A quick scan of the volume suggests that this prototype system, constructed by the National Research Council of Canada, is noteworthy for the quality of the design and its implementation rather than for any fundamental novelty. Jaksic remarks (p. 95) that they have sought "high universality, modularity, accuracy, speed, reliability, stability, flexibility and convenience in operation", a goal which is reported to have been achieved by "the optimal design and choice of components and the proper integration of the system's hardware and software". According to Blachut (p. 93) the instrument was to have been constructed by Instronics Ltd., of Stittsville, near Ottawa, under contract and close collaboration with the National Research Council of Canada. Instronics, however, was bought by Gestalt International Ltd., and Gestalt declined to build the system. The Canadian Marconi Company (CMC) of Montreal secured the license for construction, and as of the time of publication of the cited reference was proceeding with manufacture.

1.5 Research System Summaries

1.5.1 – Area-based correspondence methods

Gimel'farb Marchenko and Rybak System 1972

"An Algorithm for Automatic Identification of Identical Sections on Stereopair Photographs," G. L. Gimel'farb, V. B. Marchenko, and V. I. Rybak, Kybernetica (translations) No.2, p. 311-322, March-April 1972.

The authors of this paper were the first to use dynamic programming in a stereo correspondence process. The algorithm they describe processes image pairs on a line-by-line basis ... exploiting epipolar geometry constraints, and using known (a priori) disparity and surface slope limits to constrain the correspondence search. It optimizes a cost function of normalized cross-correlation. The convolution incorporates a lateral inhibition computation. The correspondence algorithm is described analytically as finding the function mapping intensities from one image line to the other. Testing was done on short wide images (i.e. 5x500). The authors suggest that one could improve the speed of such stereo processing in two ways. First, in using the results of prior line analyses to guide the matching and bound the search on subsequent lines, and second, in partitioning lines into smaller stretches, reducing the combinatorics of the correspondence matching. The first is a technique that CDC used in their stereo work (as will be discussed). The second can be seen as a preview of the multiple resolution correspondence processes of Baker, when it is seen that rough alignment of corresponding parts of the two lines must be made before breaking them into smaller stretches. Depiction of the results obtained with the algorithm are a bit sketchy, as the plots shown are of single line analysis only. The report says that results from this totally automated process are comparable to those of human operators using automated photomapping devices.

Levine and O'Handley System 1973

"Computer Determination of Depth Maps," Martin D. Levine, Douglas A. O'Handley, Gary M. Yagi, Computer Graphics and Image Processing, 2, p. 131-150, 1973.

This system described by Levine and O'Handley was intended to provide depth information for the Mars rover vehicle's autonomous navigation. Tests of its performance were carried out on stereo imagery collected in the vicinity of the Jet Propulsion Laboratory. Because of the system's intended use, it was possible to work with the basic premise that the scene viewed is approximately planar, running off to a horizon somewhere in the distance (not necessarily in the images). It uses collinear epipolar imaging for its two cameras to limit correspondence search. Matching is by intensity cross-correlation, with an adaptive window size set by the variance at pixel (i, j) in the image – a large variance sets a small window size, and vice versa. Processing was organized to run in lines from bottom to top. Search constraints on possible disparities are exploited throughout the analysis. First the top and bottom lines are correlated to estimate the overall disparity ranges (notice that this presupposes that scene depth varies regularly from top to bottom, as in a view toward the horizon). Then a prepass analysis is applied to a sampling of n lines ($n = 5$) to set local maximum and minimum disparity ranges. Correlation along a line pair is over windows with locally maximal variance, called 'tie-points'. The local maxima are used to iteratively segment the reference line. A coarse search using statistical parameters (variance) of image windows is used to find good candidates for the more expensive computation of the correlation coefficients. The candidate pairings chosen through this process are then evaluated to select the optimal matches and to refine their positions in three space. The coarse search is done with every other pixel along a line. Cross-correlation is only done with windows of similar variance. The system uses the epipolar geometry constraint in a way that prohibits positional reversals along a line. The authors indicate in the paper that they are aware of the difficulties introduced by occlusions, and mention an ad hoc scheme for preventing parts of the images felt to be occluded from being matched, but the technique is not further described. Two-dimensional proximity is also used to limit disparity possibilities; an allowable range is set at each tie-point by examining neighbouring disparity values on the preceding line (actually the current line minus 4 — i.e. they process every fourth row and every second element). Final disparity values are smoothed, and deviants are removed.

Mori Kidode and Asada System 1973

"An Iterative Prediction and Correction Method for Automatic Stereocomparison," Ken-Ichi Mori, Masatsugu Kidode, Haruo Asada, Computer Graphics and Image Processing, 2, p. 393-401, 1973.

Mori, Kidode and Asada, in this short paper that raises as many questions as it answers, describe an interesting stereo mapping system. Epipolar geometry is used to constrain the search for correspondences in the area-based correlation they use. The system is demonstrated on a model pattern and a pair of aerial photographs, although only a single line of results is presented. A gaussian weighted correlation function is used to diminish the effect of peripheral intensity variations. Window size is modified by the range of disparity expected for the point, and they suggest that this should be set by first correlating over a large window, then narrowing to a smaller window when the gross disparities are known (the paper doesn't explain this resolution reduction process any further). An assumption of scene continuity is also used in limiting correspondence search. The technique is iterative: the right image is repeatedly distorted and compared with the left image until no substantial intensity differences are found. The abstract says that the first matching is done on highly contrasting parts of the images ('roads, coast, forest edges'), and the context of this is used, with the smoothness assumption, to expand the correspondences into neighboring parts of the scene, but the body of the paper does not elaborate on this. The paper is very brief and cursory, suggesting much more than it reveals... it would be very interesting to see whatever further documentation they have on this system. Examples are incomplete and inconclusive. No follow-up has occurred to this work.

Hannah System 1974

"Computer Matching of Areas in Stereo Images," Hannah, Marsha Jo, Ph.D. Thesis, Stanford Artificial Inteligence Laboratory, AIM-239, Ph.D. thesis, July 1974.

This thesis describes a series of techniques developed for increasing the efficiency of area-based correlators. It contains a nice discussion of the differences between Discrete Correlation, Normalized Cross-Correlation, Normalized RMS correlation, and Normalized Absolute Difference. The work takes an exploratory approach, and documents the improvements arising from:

- correlating over a sampling of the image arrays, then refining the match estimate using the full arrays at the point having *maximum correlation coefficient* (this is referred to as 'gridding'),
- correlating over reduced resolution depictions of the images, and then refining match estimates with the higher resolution depictions,
- abstracting area characteristics (mean/variance), and using these more symbolic descriptions for limiting windows to be cross-correlated,
- using known camera geometry constraints to limit search.

A region growing approach is taken in expanding correspondences outward from matched pairs (*using an assumption of surface continuity*). Various heuristics are introduced for inferring the distinctions between occlusions, corrrespondence errors, and out-of-scene overlaps. Hannah introduced here, through the autocorrelation function, a means of assessing the quality of area-based matches.

Panton System 1978

"A Flexible Approach to Digital Stereo Mapping," Panton, Dale J., Photogrammetric Engineering and Remote Sensing, Vol.44, No.12, December 1978.

The author describes a system for obtaining a dense digital depth map of smoothly rolling terrain. The algorithm, using intensity cross-correlation, processes from left to right in the images, and so, once initialized, can use local context of previous matches and estimates based on the epipolar geometry to provide tight constraints on possible correspondeneces. Maximization of a correlation coefficient in the chosen area selects the appropriate match. About 1% of the pixels in an image are matched in this manner, although the entire image is used in determining the match correlation coefficients. Positioning accuracy of Somewhat better than one pixel is obtained. The system is able to tailor sampling window shape in one image to follow roughly the deformation of the rectangular window it matches in the other image. This window-shaping issue is one of the principal difficulties of cross-correlation analysis — only in the case of *flat terrain normal to the line* of sight are corresponding windows in the two images of the same shape. Panton's solution to this is to approximate the rectangular source window by a trapezoidal window in the other image. The technique is based on a large sampling of the *surrounding* neighborhood, and uses the terrain relief predicted by previous neighboring correspondeneces to estimate the shape of the trapezoid about a candidate surface point. This algorithm has been implemented in an experimental parallel processing *machine* which seems to achieve quite impressive performance in the processing on relatively smooth natural terrain. It is not clear whether or how much operator intervention is required.

Moravec System 1980

"Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," Moravec, Hans P., Stanford Artificial Intelligence Laboratory, AIM-340, Ph.D. thesis, September 1980.

Moravec's research was aimed at providing vehicle control information from visual sensing. His aim was not to construct a depth map, but rather to sample interesting points in a scene, and use these to provide motion calibration information and obstacle cues. There are three main vision contributions in his research: the ***interest operator***, the ***binary correlator***, and ***slider stereo***, the first two of which have been widely adopted by researchers in the field. The ***interest operator*** and ***binary correlator*** date to 1974. The ***interest operator*** is a filtering technique for selecting points at the center of locally maximal directional variance — these are typically corners. The ***binary correlator*** finds the best match of a feature in one image with the intensities in the other image using a resolution varying technique. Each feature (as found by the ***interest operator***) is represented as a series of (5) 6x6 windows, in increasing resolution (i.e. 6x6, 12x12, 24x24,.. in the original image). The lowest resolution description of the feature from the reference image is moved a pixel at a time over the other reduced image, calculating correlation coefficients at each location. The largest correlation coefficient is taken as indicating the best match. The next higher resolution window (i.e. next smaller window) centered on this is then searched (with the next higher resolution of the feature). This correlation process continues until a 6x6 patch is matched in the unreduced image. The correlation has about a 10% error rate. In ***slider stereo***, lateral movement of a camera along a track provides 9 equally spaced camera stations. Correlation of the resulting 36 (9 choose 2) possible image pairings provide a series of estimates of distances to scene points. These estimates are represented as gaussian distributions (mean equal to the distance estimate, and the standard deviation inversely proportional to the baseline) weighted by the correlation coefficient of the feature matches (from the binary correlator). The 36 histograms (distributions) are then summed, and the peak taken to indicate the correct match. Stereo tracking between vehicle motions is also performed with the ***interest operator/binary correlator*** techniques. Here, features from the central image at the previous position are searched for in the central image of the current position, and the results of this correlation inform the system of the vehicle's actual movement. The positional and depth information obtained from these correlations provide data for the navigational control of the vehicle. It knows roughly how far it has moved through the scene, and where its obstacles lie. Feature sampling is chosen so as to cover most of the scene, uniformly.

Gennery System 1980

"Modelling the Environment of an Exploring Vehicle by Means of Stereo Vision," Gennery, Donald B., Stanford Artificial Intelligence Laboratory, AIM-339, Ph.D. thesis, June 1980.

Gennery's system [Gennery 80] was designed to provide depth data for vehicular autonomous navigation. It uses cross-correlation to position points in space. The system incorporates a ground plane finder (utilizing Moravec's ***interest operator*** and ***binary correlator*** [Moravec 80]) that estimates a plane in the scene above which most points lie, and uses this to estimate the camera relative orientations. This derived camera relative orientation information enables the matching of corresponding windows to be constrained to a one-dimensional search. Having estimates of scene noise characteristics (variance, and gain and bias between the two images), he defines a correlation measure that provides sub-pixel positioning of corresponding windows. Accompanying these are estimates of the confidence and accuracy of the correspondences. Since it progresses across an image from left to right, his algorithm can use local context of previous matches to suggest tentative match sites. If these are inadequate for unambiguous matching of the particular window, search constraints based on the epipolar geometry can be used to provide further suggestions for the correspondence. These begin at the *infinity* point of the corresponding epipolar line (disparity equals zero), and come forward (to the left, with increasing disparity) until either a suitable correspondence is found or some already matched windows are encountered. When the correct locale has been chosen, maximization of a correlation coefficient in a vicinity of the selected area determines the local best match. This analysis is followed by a process of fitting ellipsoids to the determined elevation data. These, he contends, are an appropriate shape representation for use in obstacle avoidance calculations and scene matching.

1.5.2 – Feature-based correspondence methods

Arnold System 1978

"Local Context in Matching Edges for Stereo Vision",
R. David Arnold, Proceedings of the ARPA Image
Understanding Workshop, Boston, p. 65-72, May 1978.

This paper describes an edge-based stereo correspondence system which uses edge orientation and side intensity, and edge adjacencies in determining the set of globally optimal edge matches. Examples are shown of the processing of aerial views of an aircraft, cars in a parking lot and an apartment complex. The Moravec interest operator and binary correlator [Moravec 80] and a high resolution correlator and camera solver [Gennery 80] are used in determining the relative orientations of the two imaging stations. The Hueckel operator [Hueckel 71] is applied to the images, producing a set of edge elements for the correspondence. The derived camera attitude information is then used to reorient the edges to a canonic frame — one where the stereo baseline is along the x-axis and disparity shifts due to the tilt of the ground plane are cancelled. Disparities are restricted to those lying between zero (the ground) and some a priori limit in the x direction. A list of possible matches in the right image is obtained for each edge in the left image. Loose thresholds are used to specify the adjacency structure of the edges. A reinforcement/inhibition voting scheme is applied to the adjacency structure and match list, and the resulting maxima are chosen as the correct matches.

The technique uses many heuristics and thresholds, and is said to be quite sensitive to the output of the Heuckel operator.

Control Data Corporation's Automatic Planar Surface System 1979

"Automatic Stereo Recognition of Man-Made Targets," Henderson, Robert L., Walter J Miller, C. B. Grosch, Society of Photo-Optical Instrumentation Engineers, August 1979, and Henderson, R. L., "Geometric Reference Preparation Interim Report Two: The Broken Segment Matcher," RADC-TR-79-80, April 1979.

These papers describe the 'Broken Segment Matcher' phase of the Control Data Corporation's stereo research system. The aim of the work is to provide automatic reference preparation capabilities, the references being structural models of buildings which may then, at a later point, be used in scene recognition for autonomous guidance. Because of this aim, they address their work specifically toward the problem of constructing planar models of rectilinear cultural scenes from aerial imagery. They take an interesting edge and area-based approach to the problem, using edge information to guide the application of their dynamic programming intensity correlation. Roughly, their algorithm functions as follows:

- Geometrically transform a pair of images, bringing them into a *collinear* epipolar geometry (making corresponding lines parallel).

- Locate (via a Sobel operator) and 'thin' edges in the two images.

- Establish edge correspondences in the first pair of epipolar lines by hand.

- Maintain two cooperating correspondence processes to minimize the effects of image noise and extraneous detail. The first process matches intensities using only edges deemed to be 'reliable', such as those seeded to the system through the manual startup. The second process considers *all* edges, and, using the correspondences found by the first process for the particular line correlation it is presently performing, suggests a larger set of correspondences. Those correspondences which are seen to 'persist' over several preceding second process line analyses (implying that they arise from true scene geometric discontinuities) are given for consideration to the first process for its *next* line analysis.

The correlation's metric is pixel intensity difference. The two processes both use a least squares minimization on these intensity differences to choose the optimal edge correspondences. Edges are used to bound the linear regions, or intervals being correlated, and edge correspondence is a side effect of the intensity correlation - edges themselves are not compared.

The algorithm progresses from one image epipolar line to another, propagating results (to limit subsequent search) as it goes. The algorithm, as noted in the summary, requires manual starting. It propagates determined correspondences along paths of proximal edges as it progresses from line to line. Constraints have been built into the system to make it only applicable to planar surfaced structures, and the correlation only accepts transitions indicative of nearly horizontal or vertical walls ... in fact, they go to substantial effort to ignore surface detail (such as roads, sidewalks, windows). Correspondence is determined by a least-squares optimization technique. The algorithm preprocesses the imagery data in a way that precludes it from working with anything other than straight lines (as derived from sequences of edges) in the images. They have processed and documented the analysis of a single scene with their algorithm.

Their aim was to produce a three-dimensional planar rectilinear description of cultural scenes. The results shown do not indicate that they have succeeded. One point to note is that their use of two correspondence processes, with the second introducing 'new' and removing 'old' scene structures from the analysis, introduces a hysteresis into the processing — new structures (in the direction of processing) take a while to be believed ('persist'), while old structures take a while to disappear once passed. Precision would seem not to have been one of the desired properties of their system. Further, a recent paper from the group comments on the instability and 'noisy' nature of the two-process structure ([Degryse 80]), and explains several constraints they propose introducing to reduce the effects of these problems (see summary under [Panton 81]). The constraints — the scene is imaged orthographically, the structures are strictly rectilinear, all vertical surfaces are either parallel or orthogonal, and all horizontal surfaces are parallel — are severely restrictive, and have no provision for the generality and flexibility a reasonable stereo system must have. Once introduced into the analysis, it is difficult to conceive of how these restrictions could be removed for the processing of more general domains. The constraints they have used serve to bound the applicability of their process, rather than bounding its cost. Constraints should be derived from observations on *general* cases and function to distinguish realizable from impossible or unlikely interpretations. These criticisms aside, however, the CDC group's approach was fairly comprehensive, and their use of dynamic programming for the optimization is a considerable contribution (dynamic programming for stereo correspondence was first documented in a paper by Gimel'farb et al. ([Gimel'farb 72]) in 1972).

Control Data Corporation Structural Syntax Approach 1981

"Geometric Reference Studies," Panton,D. L., C. B. Grosch, D. G. DeGryse, J. Ozils, A. E. LaBonte, S. B. Kaufmann, L. Kirvida; RADC-TR-81-182, Final Technical Report, July 1981.

The authors, noting the inadequacies of the previous CDC stereo system, [Henderson 79a, Henderson 79b], calling it 'blind' to the surrounding context of the cultural scene, argued that they needed 'techniques that are more akin to human picture processing where image elements are symbolically related and where one exploits a priori knowledge of how cultural structures are constructed.' They designed a 'structural syntax ... to fulfill this geometric need.' The structural syntax is introduced as a set of geometric principles specific to the sort of 3-d cultural scene their research addressed.

This work is concerned with the processing of urban structures stereoscopically projected either orthographically or centrally to planar imaging surfaces. Applications in the report were restricted to structures in the form of right parallelepipeds. Specifically, symbolic matching was applied only to roofs, and these were restricted to horizontal and rectangular, L-shaped, or T-shaped figures. A substantial degree of operator intervention was required in the processing.

Three geometric principles were introduced as 'structural syntax':

1) The edge orientation principle: Use of convergence of 3-space parallel lines to vanishing points, for clustering parallel (opposing) edges (utilization of vanishing points has also been suggested in [Liebes 81]).

2) The principle of known transform slope: Governs allowable 3-space orientations of edges, constraining surfaces to be either vertical or horizontal.

3) The min-max transform principle: Governs the range of acceptable heights for structures.

Vanishing points are currently manually determined. The authors appear to assume that the alignment of buildings is given, so here the syntax is used to restrict the nature and orientation of walls and roofs.

Testing was done on real imagery from a small portion of one medium sized building. Edges selected were quite conspicuous, and were oriented in only one direction. (Images are *collinearized* to have epipolar lines parallel before stereo processing is begun.)

Two complementary algorithms were developed for region matching. Both centered on roof identification, under the assumption that roofs were less likely to be occluded or contain shadows. Roofs are assumed parallel to the ground. Walls are 'dropped' to intersect the ground after roofs are identified. Both systems use edge files to construct roof regions. One of the two algorithms 'constructs regions individually and separately in each image', and then performs the matching process between the images. The other first matches edges in the two images, using 3-space height as a constraint, and then constructs regions from stereo pairs of lines.

ALGORITHM 1: FIGURE MATCHING - Polygonal figures are found in the two images. Symbolic stereo matching is guided by syntax directed corner matching. Syntactic principles guide completion of incomplete line figures.

Tests have been performed on manually produced line files, using real camera model transformations. The files corresponded to horizontal surfaces. Matching proceeded satisfactorily, and incomplete figures were correctly completed.

ALGORITHM 2: LINE SEGMENT MATCHING - This procedure involves pair-wise matching of lines from stereo images. Consistency metrics are introduced to measure the validity of matches. Application has been made to rectangular and tee-shaped roofs. After rectangles and tees have been formed, they are compared, and 'line-sharing-conflict' criteria are applied to choose a final set of building tops.

The authors acknowledge 'that extensive testing and technique development is still necessary within all areas.' The present capability appears to require a substantial degree of skilled operator intervention involving iterative tuning and repeated passes through the process.

Barnard and Thompson 1979

"Disparity Analysis of Images," Stephen T. Barnard and William B. Thompson, Computer Science Department, University of Minnesota, Jan. 1979.

This paper presents an algorithm for matching points across two images, with the intent of either determining three-space position (with camera motion) or spatial velocities (for object motion). Points are chosen, tentatively matched, and then iteratively corrected through the assumption of scene continuity. Point selection uses the Moravec interest operator and a 5x5 correlation window to select locally maximal intensity variance locales. A threshold is used to choose significant variations. The correction process presumes that scene depth is continuous, and does a relaxation-type iteration to improve the confidence in similarity estimates. These similarity estimates are based on the sum of squared differences over the window (5x5). Estimates are iterated through an adjustment procedure 10 times, and the highest surviving probabilities are taken as indicating the correct correspondences (if probability is above 0.7). This is a very simple correspondence procedure. It cannot use camera attitude information to guide or limit the search, but because of this can be used for motion tracking as well as stereopsis.

Marr-Poggio System (Grimson) 1980

"Computing Shape Using a Theory of Human Stereo Vision,"
Grimson, W. E. L., Department of Mathematics, MIT, Ph.D.
thesis, June 1980.

The approach of the MIT group is in melding psychological theory and observations into a computational algorithm for stereo vision. They consider neurophysiological relevance and biological feasibility crucial aspects of their algorithm, and support the details of their approach with extensive references to the perceptual psychology literature. The algorithm, developed basically by David Marr and Tomaso Poggio [Marr 77], is an edge-based line-oriented filtering and matching process. Grimson's implementation of the stereopsis algorithm [Grimson 80] processes as follows:

- Fill 4 pairs of working arrays with zero-crossing values and orientations. The zero-crossings are found by convolving the images with 4 spatial frequency tuned band-pass filters, varying in size from 7 to 63 pixels in width.
- Set initial vergence values for the eyes in the two images (manually).
- Match zero-crossings in the paired arrays with these relative eye positions. Within paired arrays, the process decides upon acceptable matches on the basis of zero-crossing *contrast* (positive or negative) and very rough edge orientation estimates (quantized to 30 degrees, so slopes must be within approximately 60 degrees of eachother). Matches are of *positive, negative,* and *zero* disparity, relative to the vergence.
- Mark ambiguous or 'no-match' edges as such.
- Check unmatched points in *regions*, and for those where this number is greater than 30%, delete all matches. Regions are defined with regard to some statistical measure to ensure that the size represents a reasonable local sample.
- On the basis of low frequency filter matchings, make various *positive* and *negative* vergence movements to bring unmatched high frequency edges into correspondence (high frequency edges come from the smallest filters), and iterate on the matching process.

A subsequent process interpolates a smooth surface to this derived edge-based disparity data, resulting in a full depth map. The assumption that allows interpolation to take place is that 'no information is information', i.e. that the lack of edge signal in a part of the scene indicates that there are no intensity discontinuities there, and thus likely no depth discontinuities. This is a valid assumption, although rather dismissive of useful intensity information. Furthermore, the interpolation scheme makes no distinction between surface boundaries (depth boundaries) and surface details ... the former should likely be breakpoints for the interpolation, rather then knots. The resulting surface fitting smooths over the entire scene. Elegant as the interpolatory analysis is, a true solution to the problem of defining inter-edge surface shape would need to consider the global context *('is there any indication that this is an occluding edge?')* and, where possible, domain knowledge *('does this seem to be the top of a building?')*.

The results published include the analysis of several random dot stereograms, each composed of 4x4 randomly positioned black and white squares, with the maximum vergence variation running from 2 to 6 dot widths. Other examples include a ground level building scene, a view from a Mars Viking vehicle, and a random dotted coffee jar.

Assessment of the algorithm is a bit difficult: it uses a fairly simple control structure with unsophisticated matching criteria, and its success from these mechanisms is quite remarkable. But questions arise. The approach lacks a mechanism for assessing **global consistency** in its correspondence results. It would seem from the discussion of the algorithm that the initial eye vergence plays an important role in determining the final set of correspondences. By accepting high frequency channel correspondences on a *local* basis the implementation precludes other vergence matchings which could be *globally* more satisfactory (it should be noted that lower spatial frequency is **not** synonymous with *globality* — see [Julesz 76]). Notice also that the low-frequency to high-frequency control structure that is said to be as used here is shown in [Frisby 77] to be inadequate as a model for human stereopsis.). Using a maximum filter size that corresponds to the largest observed in foveal vision only (the implementation doesn't vary filter size with eccentricity, as the theory suggests), Grimson has excluded from his processing the possibility of the more globally-driven radical vergence movements that seem necessary for scenes having large disparity variations. Perhaps this would be recoverable through the correct implementation, with filter size varying with eccentricity ... he has only implemented the theory for foveal vision. **Monocular cues**, which their theory doesn't address, are known to provide information for such radical vergence movements ([Saye 75]). Initial vergence is set manually; it is not clear how subsequent major vergence adjustments are controlled. In fact, several control strategies are experimented with in the text, each to give the optimal results for the channel noise settings being tested. No clear definition of vergence control is given. In the light of the chronic failure in past vision research to document limitations and test to the breaking point, it may seem rather unfair to bring criticism to an apparently successful algorithm such as this, but its completeness has yet to be demonstrated (an interesting recent extension of the Marr-Poggio theory of stereopsis that addresses some of these issues is described in [Mayhew 81]).

Baker System 1981

"Depth from Edge and Intensity Based Stereo," Baker, H. Harlyn, University of Illinois, Ph.D. thesis, September 1981.

The depth determination system described in this thesis is founded on an edge-based line-by-line stereo correspondence scheme. Its processing consists of:

- extracting edge descriptions of a pair of images (with a simple second difference zero-crossing edge operator),
- linking these edges to their nearest neighbors to obtain the connectivity structure of the images,
- correlating the edge descriptions on the basis of local edge measures, and
- cooperatively removing those edge correspondences formed by the correlation which violate the connectivity structure of the two images.

Two further correspondence processes fill in surface detail missed by this edge mapping/removal scheme. The first is another edge-based matching that is restricted to the intervals defined by the correspondences found earlier. The intention here is to pair edges that were either unmatched or found too rivalrous by the total line matching algorithm, but are readily paired within the tighter context of corresponding intervals (although Baker mentions that this phase of the correlation is still in development). The final correspondence process is one which uses a technique similar to that used for the edges but operating on the image intensity values within corresponding intervals. The result of the processing is a full image array disparity map of the scene. Dynamic programming, in the form of the Viterbi algorithm, is used for all (there are four) correspondence processes.

The system incorporates both an edge-based and an intensity-based analysis in its processing. Edges are matched in a near-optimal line-by-line dynamic programming correlation. This correlation allows edges to remain unpaired (from which they may be considered either occluded or spurious). Edges in reduced resolution versions of the images are first correlated to provide rough estimates of the scene correspondences. These then constrain the disparity values possible for edge matchings in the full resolution images. Properties of the edges used for the matching are contrast, intensities to the sides, orientation in the image planes, and proximity to nearest (correlable) edges along the epipolar line. The use of the last two parameters is based on an analysis of Arnold and Binford ([Arnold 80]). The dynamic programming algorithm maximizes the decision metric, giving the best (local to the line) interpretation of the edges as part of a realizable scene. Two-dimensional connectivity of the edges is used to remove correspondences inconsistent with a single global interpretation. Remaining matches form a kernel of good correspondences from which the further edge-based and then intensity-based dynamic programming correlations complete the image matchings up to the point of a full disparity map.

Its exploitation of line-by-line redundancy and 2-D continuity for global consensus free this system from the need for either manual initialization (as [Panton 78], [Henderson 79a], [Grimson 80]), or sequential line processing (as [Levine 73], [Panton 78], [Henderson 79a], [Gennery 80]), thus avoiding the problems of predictive matching where miscorrelations can lead to deviation from the correct surface solution. This is the first stereo system to make explicit use of continuity along extended contours in disambiguating correspondence rivalries, and it uses a considerably more sophisticated matching metric in its correspondences than any of the above systems. Noticeable is the fact that the correlation performed on intensity values needs improvement; although its disparity context is quite well defined by the 'kernel' of good edge correspondences, the pixel intensity correspondence is subject to line-by-line aberrations. Smoothness in disparities along the line is controlled by the optimization, but no such constraints exist for *between* line disparities – a better, two-dimensional fitting algorithm is needed. Results show the processing of two quite disparate scenes, but demonstration on more imagery is necessary to convince one of the algorithm's generality.

Arnold System 1982

"Automated Stereo Perception," R. David Arnold,
Department of Computer Science, Stanford University,
forthcoming Ph.D. thesis, 1982.

This thesis deals with feature-based stereo correspondence. The use of epipolar geometry makes the matching a one-dimensional problem. Edge continuity is utilized in combining matches along adjacent epipolar lines to produce a match over an entire image.

An important component of the thesis is the derivation of two specific analytic functions. These functions are based on geometric constraints for the matching of edges in the one-dimensional stereo correspondence problem. One concerns constraints on the *intervals* between adjacent edges and the other concerns constraints on the *angles* of matched edges. These results allow a distribution function in the object space to be translated to a distribution function in the image space. The functions allow probability estimates to be made on the likelihood that edges from the two images correspond, and can be used in the selection of the best pairing among a set of alternatives. The analysis suggests that the functions are sharply peaked even for the 60 degree vergence angles used in aerial photography. When angles corresponding to human vision are used, the conditions are extremely strong. The author has modified a dynamic programming algorithm to incorporate these matching constraints with an interpretation mechanism that requires an explanation for each occluded edge or surface (interval).

A globally optimal match may be sub-optimal when viewed from the narrow context of a single epipolar line. Edge continuity across epipolar lines is used to take the one-dimensional matchings into a globally consistent two-dimensional matching. This is achieved through an extension of the Viterbi algorithm which produces for each line matching a list of all matches scoring within a preselected range of the optimal match. This list is filtered by an iterative process which enforces consistency among adjacent epipolar lines.

The system assumes an input of linked edges; that is, extended edges rather than edge elements (in fact, the current implementation is based on straight line segments). Input files were prepared by hand through computer-assisted tracing of real imagery. Effort was made to reflect the imperfection of real data. Constraints based on image intensities are used only very weakly . The output of the system is a three-dimensional map of edges in the scene.

The principal contributions of this thesis are the two probability functions and the modification to the Viterbi Algorithm enabling sub-optimal paths to be computed and retained for later disambiguation. It relies on clean edge data (in the dynamic programming optimization every edge must be explained, yet none can be considered spurious), and manual processing for the connectivity, orientation and position of edges, and the intensity values of surfaces. These restrictions make it difficult to judge the system's potential in real imagery, automatically processed. The contributions will doubtless be used in future stereo correspondence processes ([Baker 1981] exploits the probabilistic interval and edge orientation measures derived here).

1.6 References

[Allam 78] Allam, M.M., "DTM Applications in Topographic Mapping," *Photogrammetric Engineering and Remote Sensing*, vol. 44, no. 12, 1513, December 1978. (cited on p. 2,16)

[Arnold 78] Arnold, R. David, "Local Context in Matching Edges for Stereo Vision," *Proceedings of the ARPA Image Understanding Workshop*, Boston, 65-72, May 1978. (cited on p. 3,5,6)

[Arnold 80] Arnold, R.D., and T.O. Binford, "Geometric Constraints in Stereo Vision," *Soc. Photo-Optical Instr. Engineers*, vol. 238, Image Processing for Missile Guidance, 281-292, 1980. (cited on p. 28)

[Arnold 82] Arnold, R. David, "Automated Stereo Perception," Department of Computer Science, Stanford University, forthcoming Ph.D. thesis, 1982. (cited on p. 3,6,7,9)

[Baker 81b] Baker, H. Harlyn, "Depth from Edge and Intensity Based Stereo," University of Illinois, Ph.D. thesis, September 1981. (cited on p. 3,6,8,29)

[Bertram 63] Bertram, S., "Automatic Map Compilation," *Photogrammetric Engineering*, January 1963. (cited on p. 2)

[Bertram 65] Bertram, S., "The Universal Automated Map Compilation Equipment," vol. 15, part 4, International Archives of Photogrammetry, 1965; *Photogrammetric Engineering*, 1965. (cited on p. 2)

[Bertram 69] Bertram, S., "The UNAMACE and the Automatic Photomapper," *Photogrammetric Engineering vol. 35*, no. 6, 569-576 June 1969. (cited on p. 2)

[Binford 81] Binford, Thomas O., "Inferring Surfaces from Images," *Artificial Intelligence*, vol. 17(1981) 205-244, August 1981. (cited on p. 9)

[Blachut 76] Blachut, T.J., chairman, "Results of the International Orthophoto Experiment 1972-76," XIII Congress of the International Society Of Photogrammetry, Helsinki, 1976, publ. National Research Council Canada, NCR 15362, May 1976. (cited on p. 2)

[Burt 80] Burt, Peter and Bela Julesz, "A Disparity Gradient Limit for Binocular Fusion," *Science*, vol. 208, no. 9, 615-617, May 1980. (cited on p. 7)

[Degryse 80] DeGryse, Donald G. and Dale J. Panton, "Syntactic Approach to Geometric Surface Shell Determination," *Soc. Photo-Optical Instrumentation Engineers*, vol. 238, 264-272, August 1980. (cited on p. 24)

[Dowman 77] Dowman. I.J., "Developments in On Line Techniques of Photogrammetry and Digital Mapping," *Photogrammetric Record*, vol. 9, no. 49, 41-55, 1977. (cited on p. 16)

[Friedman 80] Friedman, S.J., editor, Manual of Photogrammetry, American Society of *Photogrammetry*, C. C. Slama, editor-in-chief, 1980. (cited on p. 1,16)

[Frisby 77] Frisby, John P. and John E.W. Mayhew, "Global Processes in Stereopsis: Some Comments on Ramachandran and Nelson(1976)," *Perception*, vol. 6, 195–206, 1977. (cited on p. 27)

[Gennery 80] Gennery, Donald B., "Modelling the Environment of an Exploring Vehicle by Means of Stereo Vision," Ph.D. thesis, Stanford Artificial Intelligence Laboratory, AIM–339, June 1980. (cited on p. 3,4,5,7,21,22,28)

[Gimel'farb 72] Gimel'farb, G.L., V.B. Marchenko, and V.I. Rybak, "An Algorithm for Automatic Identification of Identical Sections on Stereopair Photographs," *Kybernetica* (translations) no. 2, 311-322, March–April 1972. (cited on p. 3,24)

[Grimson 80] Grimson, W.E.L., "Computing Shape Using a Theory of Human Stereo Vision," Department of Mathematics, MIT, June 1980. (cited on p. 3,6,8,26,28)

[Hannah 74] Hannah, Marsha Jo, "Computer Matching of Areas in Stereo Images," Ph.D. thesis, Stanford Artificial Intelligence Laboratory, AIM–239, July 1974. (cited on p. 3,5)

[Helava 57] Helava, U.V., International Photogrammetric Conference on Aerial Triangluation, Ottawa, August 1957. (cited on p. 1,16)

[Henderson 79a] Henderson, Robert L., Walter J. Miller, C.B. Grosch, "Automatic Stereo Recognition of Man-Made Targets," *Society of Photo-Optical Instrumentation Engineers*, August 1979. (cited on p. 4,6,7,24,28,28)

[Henderson 79b] Henderson, R.L., "Geometric Reference Preparation Interim Report Two: The Broken Segment Matcher," RADC–TR–79–80, April 1979. (cited on p. 6,7,24)

[Hobrough 78] Hobrough, G.L. and T.B. Horbrough, "Image On-line correlation," *Bildmessung und Liftbildwessen*, vol. 46, no. 3, 79–86, 1978. (cited on p. 16)

[Hueckel 71] Hueckel, M., "An Operator which Locates Edges in Digital Pictures," *Journal of the ACM*, vol. 18, no.1, pp.113–125, January 1971. Erratum in 21, 350, 1974. (cited on p. 22)

[Julesz 76] Julesz, Bela, "Global Stereopsis: Cooperative Phenomena in Stereoscopic Depth Perception," in *Handbook of Sensory Physiology VIII*, R. Held, H. Leibovitz and H-L. Teuber, editors, Springer, Berlin, 1976. (cited on p. 27)

[Kelly 77] Kelly, R.E., P.R.H. McConnell, and S.J. Mildenberger, "The Gestalt Photomapping System," *Journal of Photogrammetric Engineering and Remote Sensing*, vol. 43, 1407, 1977. (cited on p. 2)

[Levine 73] Levine, Martin D., Douglas A. O'Handley, Gary M. Yagi, "Computer Determination of Depth Maps," *Computer Graphics and Image Processing*, 2, 131–150, 1973. (cited on p. 3,4,5,8,9,28)

[Liebes 81] Liebes Jr., S., "Geometric Constraints for Interpreting Images of Common Structural Elements: Orthogonal Trihedral Vertices," *Proceedings of the ARPA Image Understanding Workshop*, 1981. (cited on p. 24)

[Marr 77} Marr, D. and T. Poggio, "A Theory of Human Stereo Vision," MIT Artificial Intelligence Memo no. 451, November 1977. (cited on p. 6,26)

[Mayhew 81] Mayhew, John E.W. and John P. Frisby, "Computational and Psychological Studies Towards a Theory of Human Stereopsis," *Artificial Intelligence Journal*, vol. 16, 1981. (cited on p. 27)

[Moravec 80] Moravec, Hans P., "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," Stanford Artificial Intelligence Laboratory, AIM-340, Ph.D. thesis, September 1980. (cited on p. 3,4,7,9,22)

[Mori 73] Ken-Ichi Mori, Masatsugu Kidode, Haruo Asada, "An Iterative Prediction and Correction Method for Automatic Stereocomparison," *Computer Graphics and Image Processing*, 2, 393–401, 1973. (cited on p. 3,4,9)

[Nishihara 81] Nishihara, H.K., and N.G. Larson, "Towards a Real Time Implementation of the Marr and Poggio Stereo Matcher," *Proceedings of the ARPA Image Understanding Workshop*, 114–120, May 1981. (cited on p. 6)

[Panton 78] Panton, Dale J., "A Flexible Approach to Digital Stereo Mapping," *Photogrammetric Engineering and Remote Sensing*, vol. 44, no. 12, 1499–1512, December 1978. (cited on p. 3,4,7,16,28,28)

[Panton 81] Panton,D.L., C.B. Grosch, D.G. DeGryse, J. Ozils, A.E. LaBonte, S.B. Kaufmann, L. Kirvida, "Geometric Reference Studies," RADC-TR-81-182, Final Technical Report, July 1981. (cited on p. 6,7,9,24)

[Ryan 79] Ryan, T.W., R.T. Gray, and B.R. Hunt, "Prediction of Correlation Errors in Stereo-Pair Images," SIE/DIAL-79-002. (cited on p. 1)

[Ryan 80] Ryan, Thomas W., and B.R. Hunt, "The Prediction of Accuracy in Digital Cross-Correlation of Stereo-Pair Images," *Soc. Photo-Optical Instr. Engineers*, vol. 219, Electro-Optical Technology for Autonomous Vehicles, 1980. (cited on p. 1)

[Saye 75] Saye, Ann and John P. Frisby, "The Role of Monocularly Conspicuous Features in Facilitating Stereopsis from Random-Dot Stereograms," *Perception*, vol. 4, 159–171, 1975. (cited on p. 27)

[Scarano 76] Scarano, Frank A., "A Digital Elevation Data Collection System," *Photogrammetric Engineering and Remote Sensing*, vol. 42, no. 4, 489, April 1976. (cited on p. 2)

[Schumer 79] Schumer, Robert A., "Mechanisms in Human Stereopsis," Ph.D. thesis, Department of Psychology, Stanford University, 1979. (cited on p. 8)

Chapter 2

SURVEY OF MODEL-BASED IMAGE ANALYSIS SYSTEMS

2.1 Introduction

This chapter makes a survey and critique of the state of the art in model-based image analysis systems. It includes summaries of a selection of systems and evaluates them from the viewpoint of progress toward general vision systems. The chapter also describes principles of design of general vision systems. Those principles are personal, dogmatic, and subject to question, but most of them have been the basis for the author's work over the last 13 years or so.

Humans pursue many goals in an unconstrained, constantly changing world with many objects. That is an example of a general vision system. Robots work in manufacturing with a single, repeated task with few objects, all known, in constrained environments engineered to simplify those tasks. Vision systems for manufacturing can use many special case tricks which don't generalize to worlds with many objects and many goals. However those systems must eventually come to resemble general vision systems more than is generally acknowledged. Consider what is required to make improvements to current industrial vision systems to make them reach a large market. A single system must be easily programmed to accomplish many tasks of visual control or inspection, without extensive special case engineering, insensitive to variation of lighting. A single system must be instructed effectively to determine many classes of defects, including rare defects and those not encountered in training, and to distinguish cosmetic *from real defects*. While a single industrial task is very constrained, the range of tasks and objects is very large across a major company, an industry or industries. It is not cost effective to build a separate system for each task, thus an adequate system must have modules from which to select to specialize to a particular task. Other applications which are now being pursued impose even more severe requirements for high speed object classification in complex environments.

We work to make machines see as men do, to understand how men see, to implement performance systems for vision applications, and to incorporate current capabilities in an experimental system to support further research. Those activities converge in the sense that they depend on common visual operations and common representations. We use the following terminology: **Prediction** refers to mappings from object models to their predicted appearances in images, i.e. top-down mappings in the direction from symbol to signal. **Description** and **Observation** refer to mappings from sensed images to perceived surfaces and objects, i.e. bottom-up mappings, from signal to symbol. **Interpretation** refers to mappings between predictions and descriptions. **Generic** means defined over a class, e.g. generic with respect to viewpoint meaning defined over a range of viewpoints.

An example of interpretation is template matching, i.e. finding the best embedding of a template subimage to an observed image, over all translations and rotations. An example of prediction is mapping a specific object model to a specific image from a particular viewpoint by techniques of computer graphics. Neither example is interesting for general vision systems.

In the bad old days, work on vision systems was commonly justified by statements that general segmentation and description were impossible, a dead-end. Description modules might be improved a bit but they could never be much better. It was not necessary to improve them. Within that doctrine, powerful vision systems could be made by combining existing modules into systems which used extensive world knowledge. Not only was this the way to successful applications programs but it was a basis for human visual performance.

A minority held a contrary view, that low level vision modules were weak, loosely speaking less than 1% of their potential. The problems were fundamental, but segmentation and description are powerful in humans and would eventually be powerful in machines. Combining ineffective procedures would produce little of interest; i.e., three pieces of junk are usually worth less than one, i.e. more to get rid of. Vision systems would be inherently limited by performance of segmentation modules. Further, ad hoc building of systems would be frustrated because without effective representation and description, knowledge cannot be used effectively. There are deep problems in encoding and using world knowledge in a general way in a wide range of situations, i.e. in relating world knowledge to image structures where world knowledge is typically about object classes in three space. From this point of view, combination of existing modules to build vision systems would not lead to powerful vision systems. Systems might be made to get by for some useful but limited applications. Building systems is also a useful exercise to look at the full problem. However, careful theoretical analysis and implementation of individual vision operations was essential as a basis for building vision systems and especially for effective applications.

Majority opinion has swung to studies of fundamental vision science. System building is no longer an end in itself, but finds motivation in applications and demonstrations. Indeed, it is time to look at applications and time to look at the total problem.

2.2 Observations About Systems Surveyed

In the survey, four systems are discussed first. Others are discussed in alphabetic order. One vision system has succeeded in labeling large regions in aerial photographs [Nagao 78]; another has labeled large regions in a few outdoor scenes from ground level views [Ohta 80]. ACRONYM has identified aircraft in aerial photographs [Brooks 81]. Another system has labeled objects in typical scenes of a desk [Shirai 78]. The results of these and other efforts are encouraging as first demonstrations. As general vision systems, they have a long way to go.

With several of the summaries of systems are comments about the limitations of that system as a general purpose vision system. The lists of limitations are only reminders, containing some examples, not complete descriptions of the ways in which these systems would not generalize. A complete description of problems is left to a section which follows summaries of systems.

What works in the best of these systems? What makes them go? They succeed with quasi-2-D scenes, for example aerial photographs, industrial scenes from a fixed viewpoint, x-ray images, and ground level photos from a fixed viewpoint. Even ACRONYM, which incorporates viewpoint-insensitive mechanisms, has been demonstrated only on aerial images, although there is reason to believe that ACRONYM will succeed with ground level photographs also. However, [Ohta 80] demonstrate some success with a set of similar scenes but from several viewpoints. Still, the system uses relations which are manifestly viewpoint-dependent.

The systems use image models. In [Ohta 80], the model includes the following relations: the sky touches the upper edge of the picture; the road touches the lower edge of the picture; trees are in the middle of the picture; buildings have a linear upper boundary with the sky. One type of image model includes maps and symbolic sketches. Several systems use maps in registration of images with symbolic data bases. An exception, ACRONYM [Brooks 81], generates viewpoint-insensitive image models from object models in three space.

The systems know few objects. [Ohta 80] have four objects with sub-objects. [Nagao 78] mention about nine classes of areas but the system described apparently distinguishes only four: fields (with vegetation or bare); buildings; wooded areas; and linear features (roads, rivers, rail). ACRONYM has models of generic wide-bodied passenger aircraft, Lockheed L-1011, Boeing 747-B and 747-SP. [Shirai 78] include a few objects found on desks: a lamp, a book stand, a cup, a telephone, and some small objects, e.g. a pen. These systems take advantage of the limited set of objects by using predominantly top-down interpretation of images, relying heavily on prediction. The systems use models of specific objects, e.g. [Ohta 80] use a model in which a car is dark with respect to the road against which it appears.

Depth information is powerful. With depth data, input and models are at the same level. Depth data directly describes surfaces, which are natural for object interpretation. Images are two-dimensional, while objects are three-dimensional. Image level invariants are weak, while three dimensional invariants are strong and valuable for object-level interpretations. The projection process destroys information. The remark is often made that vision is the inverse of graphics, as if to imply that vision is thus somehow simpler than we find it to be. The inverse of projection is indeed simple, but it is useless. There is no unique inverse, only very high orders continuous infinities of projectively equivalent surfaces. Depth can be inferred from sequences of images (stereo, observer motion, object motion, photometric stereo); a primary problem is determining corresponding elements in the sequence. Direct depth measurement avoids the correspondence problems of reconstructing depth data from image sequences. There are much greater problems in inferring surface structure and depth relations from single images.

One key to performance of systems is their use of measurements which are simply related to characteristic properties of surfaces and to three space relations of surfaces. Surface reflectivity and surface material are often characteristic, e.g. concrete and asphalt roads, or vegetation. Surface material can be partially inferred from color and texture. Intrinsic properties such as color, texture, and shape thus provide powerful and simple clues to interpretation. In [Rubin 78], 44% identification of pixels was achieved with the intensity of the blue channel alone. Remote sensing has depended upon pixel classification based on spectral properties. [Nagao 78] use spectral properties of vegetation, shape descriptors including elongation and straight boundary, and a simple texture descriptor. [Ohta 80] use color, simple texture, straight boundary, and hole, in addition to non-intrinsic relations which are viewpoint-dependent. ACRONYM [Brooks 81] utilizes shape descriptors which are ribbons and ellipses, and describes a larger class of quasi-invariant observables of shape. The use of simple intrinsic properties can be pushed much further than has been done thus far.

2.2.1 – Limitations

What are the limitations of these systems for the restricted domains for which they were designed? What are the limitations of the systems in attempting to extend them to general vision systems? There is a utility to having special case mechanisms, however it is essential to have effective mechanisms to cover expected cases, or to have general mechanisms.

In this paragraph are some personal opinions summarizing the state of vision systems. Most systems have not attempted to be general vision systems. The performance they achieve is based on a severely limited context with low ambitions for the quality of scene description they generate. ACRONYM does demonstrate some progress toward the goal of a general vision system, which is still distant. The fundamentalist view of systems appears to have been accurate: first, performance of vision systems is strongly limited by performance of segmentation modules; second, systems make weak use of world knowledge. Existing systems have weak description with little use of shape. Systems, like description operations, have achieved less than 1% of their potential, even allowing for weak description. They use little information in weak ways. Systems primarily relate image relations to image observables; they mostly lack the ability to relate three space models to images. Systems are being built, but there is relatively little emphasis on basic vision problems in system building. Until recently, systems efforts have been small and short-lived, a few man years effort. Focussed and continuous efforts are necessary but not sufficient for system building. Just the system programming effort of building a vision system is enormous.

With the exception of ACRONYM, systems surveyed do not appear to extend to viewpoint-insensitive interpretation. Since the systems depend on image models and relations, they are strongly viewpoint-dependent. To generalize would require three-dimensional modeling and interpretation as in ACRONYM.

The systems do not appear to extend to environments with many objects. They jump to conclusions based on flimsy evidence which would probably not distinguish many objects in a complex visual environment. Humans may occasionally halucinate in the same way, but usually they have strong evidence for interpretation.

The primary limitation in seeking stronger evidence for interpretation is the weak segmentation capability and weak implementation of observation primitives. For example, In [Ohta 80], shadows are identified as having low intensity; black objects would be confused with shadows in typical scenes. In indoor scenes, contrast across shadow boundaries can be low. Most systems surveyed segmented images based on pointwise properties, i.e. connected components of points which independently fall in some spectral band. Few systems used shape in segmentation of regions, e.g. used continuity of edges forming ribbons. In other words, most systems segment regions then describe their shape, while a few segment well-formed regions. Simple texture measures are used, however segmentation of extended edges is a major void, especially segmentation of textured regions. Correspondingly, shape description is primitive. Interpretation systems do not appear yet to make full use of even these limited capabilities. No system surveyed has effective texture segmentation and description to make use of surface texture as an intrinsic property in identification. Where range measurements are available, surfaces can be segmented directly at occlusions, i.e. position discontinuities. This is not the same as segmenting objects as humans do, because an object resting on a plane will not be segmented from the plane, i.e. they are coincident. Even with range data, systems make poor symbolic, segmented descriptions of surfaces, because of the fundamental weakness in segmentation.

None of the systems duplicate the human capability of color constancy, i.e. insensitivity to source spectrum. Brightness constancy is related, inferring which surfaces are white and black in black and white images. Color constancy and brightness constancy are obtained by estimating surface reflectivity by a global partial ordering of local relations [Land 77]. It is no accident that multi-spectral data is useful since surface reflectivity is characteristic in many cases (but not unique). It is also no accident that color constancy and brightness constancy have not been integrated in the systems surveyed, since constancy requires a global computation which is much more complex than pointwise multi-spectral calculation, and which requires effective segmentation. [Horn 73] has demonstrated partial success with color constancy, however that capability is not integrated in any system. To our knowledge, no use has been made of surface reflectivity in black and white images, for those same resons. In the case of an active ranging sensor, reflectivity can be computed directly at points from reflected intensity and range.

Systems typically use the hypothesis-verification paradigm. Hypothesis generation is the crucial part. Hypothesis generation is trivialized in the top-down case. For example, ACRONYM now searches only for aircraft. Another approach to hypothesis generation is evident in the systems surveyed. If an object class has a stable property, then assign the object class as an interpretation for any image region which has that property. For example, assign vegetation as an interpretation to any image region which has the right color of green. That has utility in restricted scene domains. But, in typical cases, assignment must be generalized. For example, a house has straight lines, but to assign a house interpretation to image regions with straight lines would be worth little in general. The appropriate interpretation is not necessarily the union of all objects with straight edges, since that class is very large, and the simple prescription does not tell how to break down the class. Instead, relate straight lines to structural elements with straight edges. That requires powerful capabilities for inference of shape such as those being developed in ACRONYM [Binford 81]. Thus that method of hypothesis generation is successful in restricted scene domains with a few object classes which are easily separated by local operations.

2.2.2 – Nagao, Matsuyama, and Ikeda 1978

In [Nagao 78, Nagao 80], an aerial photograph is first segmented into regions by several processes. Judging from the dominant features of each extracted region, specialized feature extraction and recognition programs are applied to that region only. Properties of objects are summarized and are fed back in order to re-analyze ambiguous regions.

To extract major regions, the following operations are used:

1. edge preserving smoothing;
2. segmentation into regions which are continuous in spectral properties;
3. extraction of five kinds of regions, called cue regions;
4. analysis of each cue region by an object detection program specific to region type;
5. summary of properties of regions fed back to subsystems.

Cue regions include: large homogeneous regions, elongated regions, shadow and shadow-making regions, vegetation regions, water regions, and high contrast texture regions. Each kind of cue region is extracted independently of the others. Some regions may appear in several types of cue regions.

The book shows data from several pictures with 4 spectrals bands, taken at low altititude from an airplane. The pictures are 256x256 with 8 bits per pixel, corresponding to 50cmx50cm on the ground. Spectral bands are red, green, blue, and infrared.

Pictures are first processed by an edge-preserving smoothing operator. It is intended to remove noise and to remove the blurring at edges of regions, for pixels which overlap two regions. It has some success; however it erodes thin lines and small regions (smaller than 3 by 3). It presumably rounds corners.

After smoothing, each picture in the four bands is divided into small patches with constant grey level. Patches with similar spectral properties are grouped together: if the differences in four bands are less than a threshold between a labeled pixel and its neighbor, the neighbor is merged. This is a gradient operation which allows shading. The threshold value in each spectral band is adaptively determined as follows: divide the differentiated picture into sixteen 64x64 blocks and make a histogram of nearest neighbor differences for each block; find a valley for each histogram where the valley has a value lower than the succeeding nine values; choose the minimum value for valleys among the sixteen blocks. This follows the reasoning that noise will have a large peak and edges will have a smaller peak.

Small regions with less than four pixels are merged with neighboring large regions with the most similar spectral properties. This is done because boundaries of homogeneous regions sometimes are ragged on account of separate smoothing in each spectral band.

One shape descriptor is the best minimum bounding rectangle (MBR), defined as the MBR with maximum ratio of (region area)/(area of MBR) over angles 10 degrees apart. Elongation and direction are taken from the best MBR.

Large homogeneous regions are defined by making a histogram of homogeneous region size and applying a valley detection algorithm. Small regions are assumed to result from noise. Regions larger than the threshold are processed further. They may be fields, grasslands, lakes, and sea.

Some regions are flat, e.g. fields and the sea; however, houses, buildings, and trees have height. Shadows give information about height; shadows are usually available because aerial pictures are usually taken in good weather. Shadows are obtained in the following way: Make a histogram of brightness of smoothed pictures. Calculate the average brightness in the whole picture. Calculate the brightness which makes the inter-class scatter minimum when divided into two classes. If the gradient at this value is small, choose it as the threshold brightness, I1. Otherwise search for a valley near that value as I1. Homogeneous regions whose average brightness is less than the value I1 are chosen as shadows. Choose shadow making regions as regions adjacent to shadows with a long common boundary in the direction away from the sun. Shadows are used to discriminate between flat objects and those with height.

Elongated objects include roads, rivers, and railroad lines. No analysis is made of rail lines. Elongated regions may be broken by cars, trains, or shadows, and they may be curved. For curved regions, the MBR does not give a good estimate of elongation; the system determines an elongation effective for curved regions by taking the longest path on the skeleton of a region. If length/width > 3, regions are called elongated.

Vegetation areas have small ratio of red to infrared intensity, a property which is quite stable. However, blue roofs have the same property. Thus, the system was made to exclude regions with large intensity in the blue band. Adjacent vegetation regions are merged into a large vegetation region. Water regions were identified by spectral properties, with the additional condition that water regions were darker than adjacent regions. Problems were found for water in shadow and where there were weeds. Extraction of vegetation and water are two examples of segmentation according to intrinsic properties of materials.

High contrast texture regions are treated as follows. After smoothing, only coarse texture remains from objects 1.5m on a side. Woods and residential areas contain small objects such as trees, houses, roads, and shadows. The authors found it almost impossible to recognize these small objects using only individual properties such as shape. Thus, the system extracts a set of small regions as a whole and recognizes each constituent region based on properties of the group of small regions. The system extracts homogeneous regions, then moves an NxN window over the image; if the window contains more than 2N boundary points, the system considers the central point of the window as part of a high contrast texture region. The system removes small holes and peninsulas by growing regions two steps, shrinking four steps, then growing two steps. The system merges any homogeneous region which has more than half its area in high contrast, but which is not a large homogeneous region. The common area between the high contrast regions and the large vegetation regions is registered as high contrast vegetation regions.

A large homogeneous region may be a crop field, bare soil, or a grassy area. Boundaries of large homogeneous regions composed of straight lines are designated crop field unless they are elongated regions. Boundaries are called straight if more than 60% of pairs of forward and backward boundary chords have angles which are less than 22.5 degrees (for point i, the forward chord is the pair of points indexed by i, i+5, the backward chord is the pair indexed by i, i-5). Crop fields so designated are put into two classes, one with vegetation, the other without. A region adjacent to a crop field or bare crop field is a candidate for either if its area is greater than half the threshold for large homogeneous regions.

Elongated regions include rivers and roads, which are used to register images. Elongated regions may be broken by bridges, cars, and shadows. They are not vegetation regions, not shadows or shadow-making regions. Road candidates are those which are elongated, not vegetation and not water regions. The system examines all pairs of candidate regions and connects those with nearly the same intensity and color, with ends which are adjacent, and with the same width and same direction. The conditions are: difference of average gray scale in the four spectral bands less than a threshold (separate for each band); ratio of widths near 1 (between 2/3 and 3/2); smallest separation between ends less than 3W, where W is the smaller of the two widths; direction difference less than 45 degrees. Roads are those with elongation > 8 from length/width along the skeleton, and for which the variance of widths along the skeleton is small. Intervening gaps may not be vegetation or water regions. The system traces any such road and picks up nearby side roads, which are connected to it, with small difference in hue, with ratio of widths between 1/2 and 2/1, and with one end point within 3W of the first road found. The system recurses on side roads to find their side roads. Cars are recognized as rectangular regions on roads.

River candidate regions are water regions, not necessarily elongated. Analysis of rivers is similar to that for roads. Multispectral properties of water in shadow are greatly affected. Thus, shadow regions are merged with water if they are adjacent to a water region, if they are not a vegetation region, and if merged regions have increased elongation.

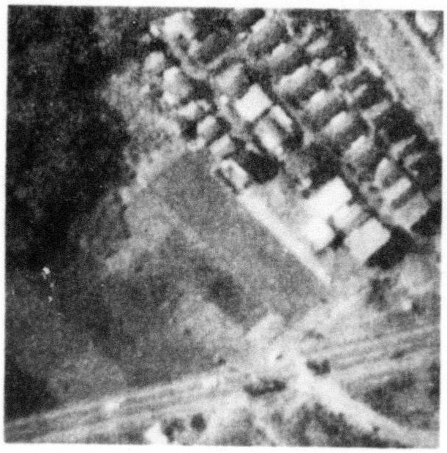
Each high contrast vegetation area is assumed to be a wood area. Small elements should have irregular shape (from the straightness condition above). A high contrast vegetation region contains shadows and shadow-making regions; this distinguishes woods from grass.

The approach for finding houses is to first identify candidates for residential areas, then find houses. High contrast regions which are not large homogeneous regions and not large vegetation areas are candidates for residential areas. Residential areas are identified as areas in which gradients are strong in two orthogonal directions. Houses typically have walls at right angles. The method would be valid where houses are laid out on an orthogonal grid. HOUSE1, a recognition routine for houses, uses these properties: in a residential area; not a vegetation region, not a shadow region, not a water region, but a shadow-making region; rectangular shape. HOUSE2 extracts regions whose average gray levels in the four spectral bands are similar to those of any houses recognized by HOUSE1. HOUSE2 requires a weaker condition on rectangularity, and waives the condition that the region be a shadow-making region if it is in a residential area. Houses also appear as two rectangular elements, where two inclined surfaces are segmented as separate, adjacent regions.

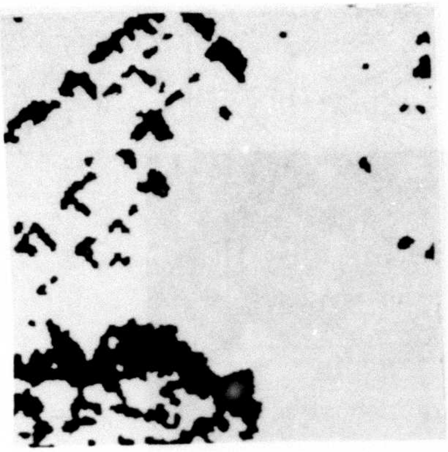
HOUSE3 tries to recognize parts of a roof which are not recognized and which are adjacent to roofs which are already recognized. HOUSE4 looks for missing houses in regular arrays. It looks for regions which fall on sites determined by 'regularity vectors' of the pattern of houses. Candidates must not be vegetation region, water region, shadow region, or elongated region, and must satisfy rectangular fit. Buildings are designated as regions which are not vegetation or water regions, which are shadow-making regions, which have area greater than an ad-hoc value, and with predominantly straight boundary.

System control tries to resolve conflict labels and to deal with unlabeled regions. If a region has more than one label, the most reliable label is chosen and other interpretations are rejected. Interpretations which are dependent on rejected interpretations (e.g. car identified from road) are also rejected. If regions are unlabeled because their shapes do not satisfy conditions of object classes, the system activates a split and merge process to attempt to split the region into two regions or to merge the region with adjacent regions. This sometimes corrects faulty segmentations. Splitting takes place at bottlenecks in width along the longest path on the skeleton of the region.

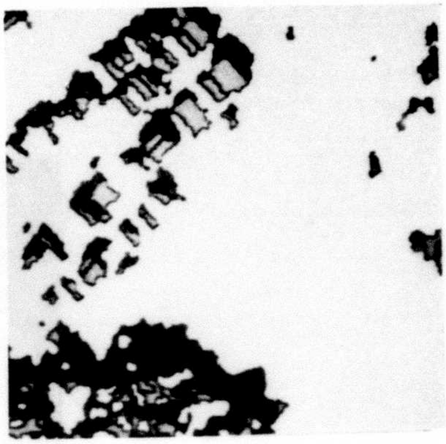
Results are shown for several images in the book. The system does well for roads, fields, and forest. Figures 2-1 through 2-24 show steps in an example of analysis (from [Nagao 78]). Some areas are falsely labeled as houses by HOUSE2 (seven in one picture). About 3% of the scene is shadow which is unlabeled. Small shadows on roads and rivers can be correctly labeled; shadowed vegetation areas can also be labeled. Otherwise, there is difficulty with large shadowed areas. A large area around houses is left unlabeled The system has difficulties with urban scenes. For five scenes, unlabeled shadow and unrecognized areas are: (3%, 19%), (1.6%, 16%), (10%, 31%), (22%, 30%), (4%, 16%). Thus 20% to 52% of area are unlabelled. This is a pessimistic evaluation since significance of many details is not related to size. The interpretation process requires about 200 seconds per multi-spectral picture on a machine with 90 nsec average instruction time, plus about 240 seconds of smoothing.



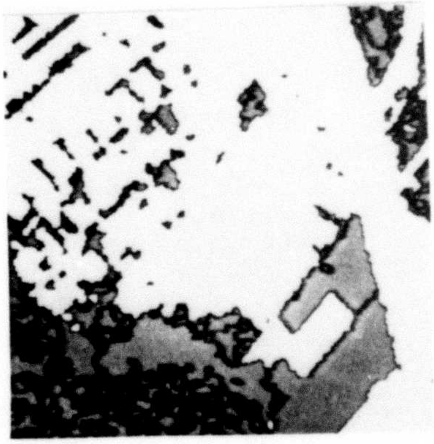
An aerial photograph (from [Nagao 80])
Figure 2-1



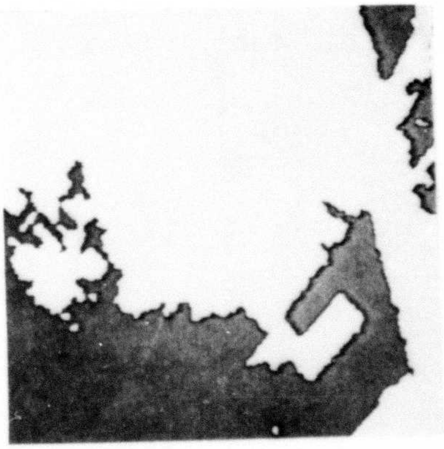
Shadow regions
Figure 2-2



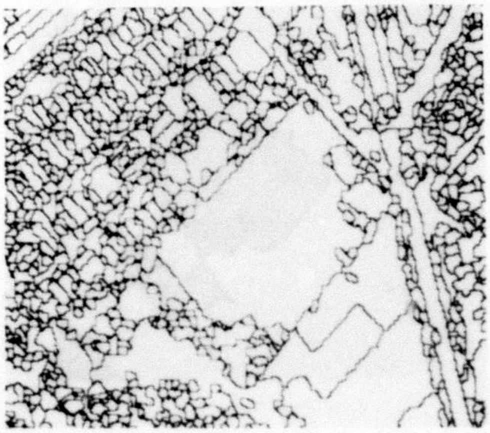
Shadow-making regions; the regions enclosed by black lines show the shadow-making regions and those shaded grey denote the shadow regions
Figure 2-3



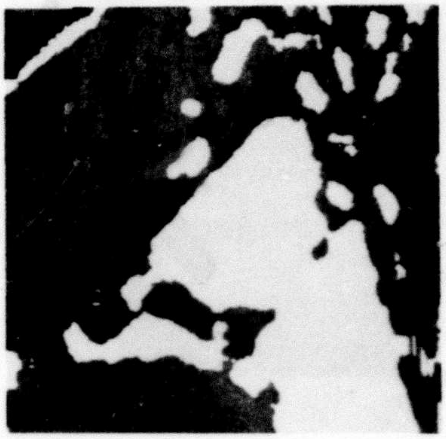
Vegetation regions
Figure 2-4



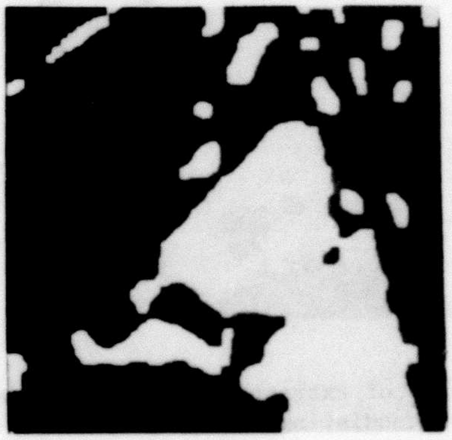
Large vegetation areas; each area consists of many elementary regions
Figure 2-5



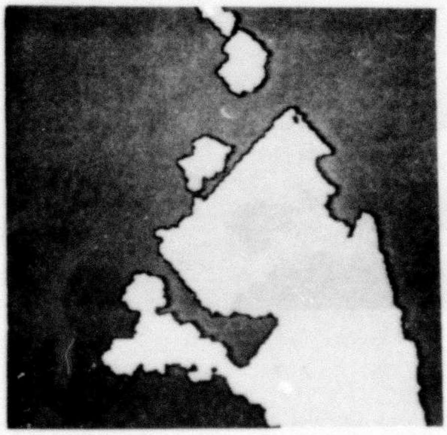
Boundaries of elementary regions
Figure 2-6



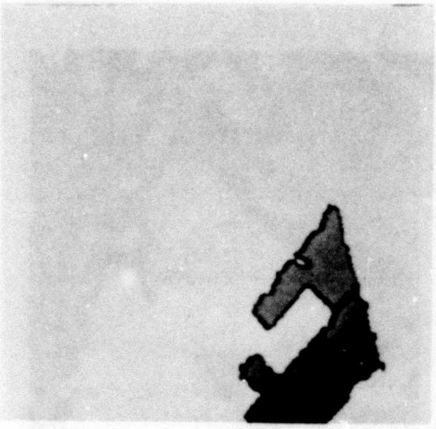
The areas with high density of boundary points
Figure 2-7



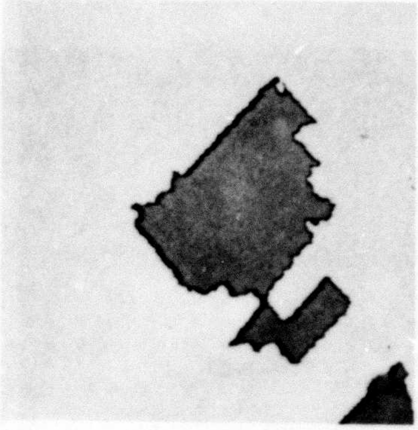
esult of removing small holes and hin peninsulas
Figure 2-8



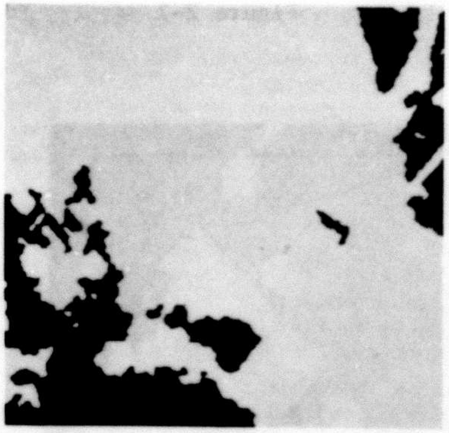
High-contrast texture area; this area consists of elementary regions
Figure 2-9



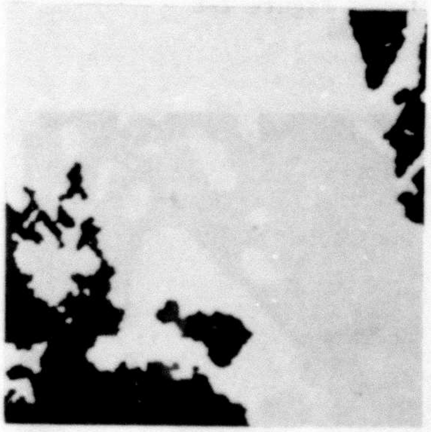
Recognized crop fields
Figure 2-10



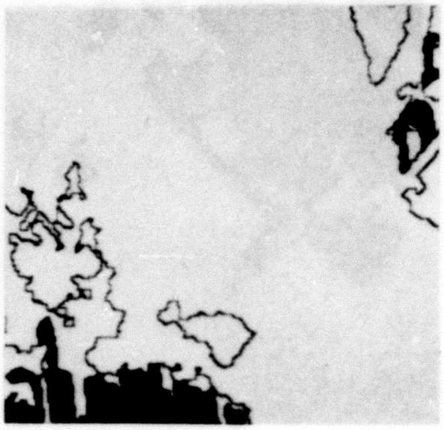
Recognized bare soil fields
Figure 2-11



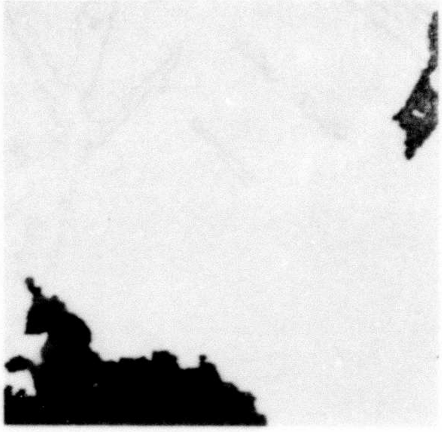
Candidate areas for forests
Figure 2-12



Result of extracting large connected area consisting of a set of elementary regions
Figure 2-13

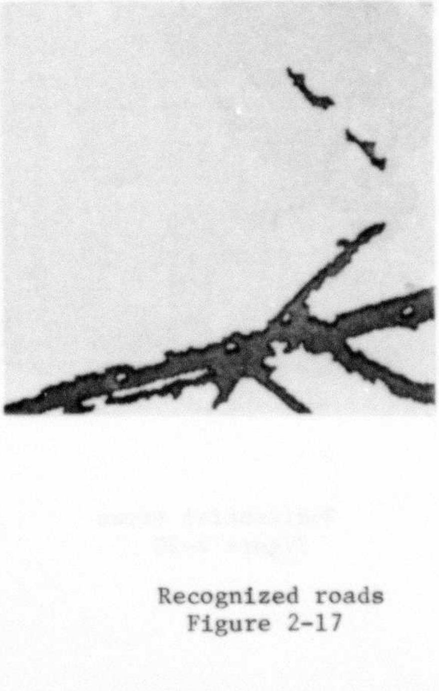


Result of extending shadow-making regions in candidate areas
Figure 2-14

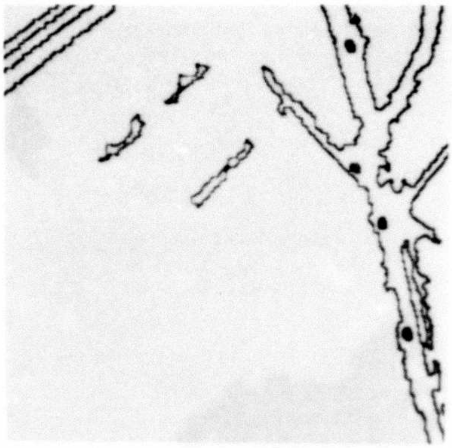


Recognized forest areas
Figure 2-15

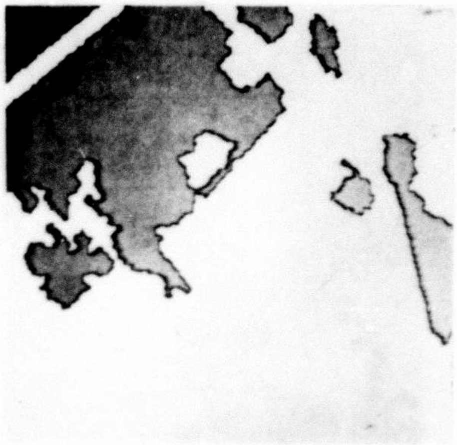
Recognized grasslands; there are some regions which are recognized as crop field and forest area as well as grassland. These conflicts are resolved by the system
Figure 2-16



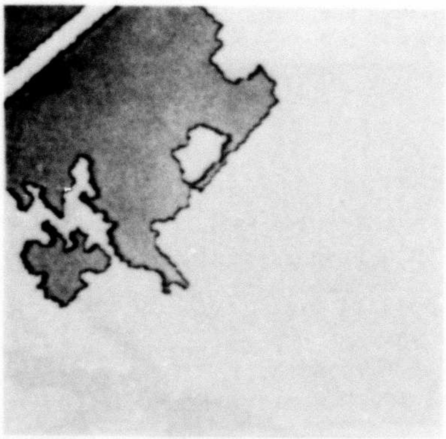
Recognized roads
Figure 2-17



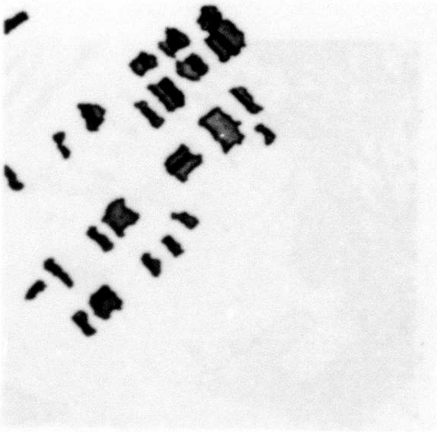
Recognized cars
Figure 2-18



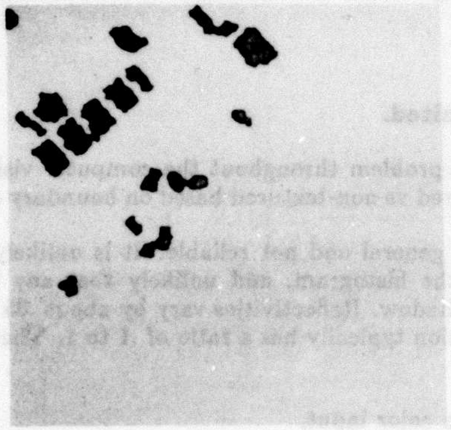
Candidate areas for residential area
Figure 2-19



Residential areas
Figure 2-20

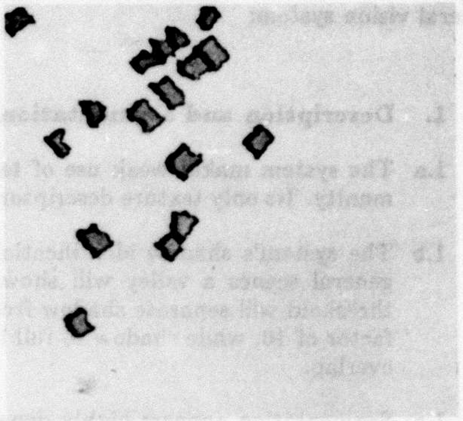


Houses recognized by the HOUSE1 subsystem; since the conditions are very strict, about half the houses are left unrecognized
Figure 2-21



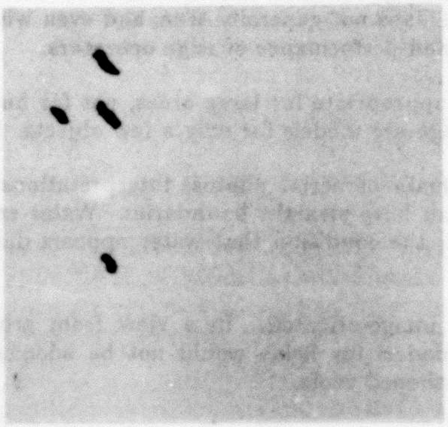
Houses newly recognized by the HOUSE2 subsystem using the multispectral properties of the already recognized houses

Figure 2-22



Houses newly recognized by HOUSE3 subsystem; two adjacent roofs of a house remerged into one region

Figure 2-23



Houses newly recognized by the HOUSE4 subsystem; "missing houses" in the large residential area are correctly recognized

Figure 2-24

This is a fine and well-crafted system. It performs interesting interpretations on these examples. Its approach is to use special subsystems to recognize specific objects. Thus, it is not

intended as a general vision system. The following limitations appear when considering it as a general vision system:

1. **Description and segmentation are limited.**

1.a The system makes weak use of texture, a problem throughout the computer vision community. Its only texture descriptor is textured vs non-textured based on boundary density.

1.b The system's shadow identification is not general and not reliable. It is unlikely that in general scenes a valley will show up in the histogram, and unlikely that any intensity threshold will separate shadow from non-shadow. Reflectivities vary by about .05 to .90, a factor of 16, while shadow to full illumination typically has a ratio of .1 to 1. Their ranges overlap.

1.c Segmentation appears highly dependent on color input.

1.d The smoothing operation degrades the picture considerably.

2. **Interpretation depends on assumptions which are not broadly useful.**

2.a Shadows provide the only three-dimensional interpretation. No use is made of shadows to determine shape of objects other than non-flat. Shadows are assumed adjacent to shadow-casting regions. This is not generally true, and even when true, makes assumptions about surface marking and performance of edge operators.

2.b Interpretation is appropriate for large areas, not for human-scale objects for which shape is important. There are models for only a few objects. Weak use is made of shape.

2.c Even for the domain of aerial photos, interpretations are made on weak assumptions. Grass pastures can have straight boundaries. Water can appear brighter or darker than surrounding land; the condition that water appears dark relative to adjacent land is not general.

2.d Interpretation is image-oriented. In a view from ground-level, fields would not be as prominent and models for fields would not be adequate. Houses would not appear as rectangular or L-shaped roofs.

2.2.3 – Ohta 1980

[Ohta 80] describes a system which assigns semantic labels to regions in color images of outdoor scenes. They present a new set of color parameters used in an Ohlander-like region analysis system which forms regions by splitting, using thresholds selected from histograms of the new color parameters. A plan is generated by an initial bottom-up coarse region segmentation. A symbolic description of the scene is made by top-down analysis using a production system with knowledge of the world represented as a set of rules.

Region analysis tends to 'over-segment', i.e. to split semantic regions (images of surfaces or objects) into several regions. In part, this is because the authors' choice of color parameters is intensity dependent. Matching searches for a many-to-one correspondence between regions and images of surfaces. Each region is evaluated for each surface interpretation rule.

[Ohta 80] determine the best set of parameters for segmentation of color regions. They propose as color parameters the three eigenvectors of the covariance matrix (Karhunen-Loeve transformation). They start out to find eigenvectors dynamically, however they find it about as satisfactory to determine eigenvectors once and for all. The eigenvectors turn out to be:

$$\begin{aligned} x_1 &=(r+g+b)/3; \\ x_2 &= \pm\,(r-b)/2; \\ x_3 &=(-r+2b-b)/4. \end{aligned}$$

Eight scenes were used in an experimental analysis. They used 109 selected regions, large regions which split into 'not-small' regions. In the wr, wb plane, eigenvector x_1 is in the first quadrant (83 regions), x_2 is in 2nd and fourth quadrants (22 regions), x_3 is in the third quadrant (4 regions). Thus, x_1 is by far the most important, x_2 next, and x_3 almost negligible. If they synthesize images based on only x_1, x_2 with a constant value for x_3, the results are reasonably good except for several small regions. That is, color is roughly two dimensional. Note that x_1, x_2, and x_3 are intensity dependent, which may not be acceptable to everyone. Regions tend not to have constant intensity, thus regions are broken into bands using this set of color parameters.

The system processes images 256x256 with 5 bits or 6 bits. Region segmentation fails with texture, so they segment off textured regions, obtained by the following: in a 9x9 window, if the Laplacian of 8 out of 9 subwindows (3x3) exceeds threshold it is considered a texture window. In a building scene, this process obtained the outer portions of a tree. The remainder (not strongly textured) is segmented by recursively applying thresholds determined from peaks of histograms of color parameters. A score is calculated for each peak, including the relative depth of the valleys and the sharpness of the peak. Regions thus obtained are evaluated by a looseness criterion, related to the fraction of border cells to total cells. The segmentation with minimum looseness is chosen. Regions with size greater than a threshold are scanned with a 32x32 window. The data structure includes regions, boundaries, and vertices. Boundaries are 4 neighbor connected, segmented into straight lines with iterative end point fits.

Only primary features are in the data structure.

1) Regions are represented by: area; mean intensity of R, G, B; degree of texture; contour length; center of mass; number of holes; scatter matrix of pixel positions; minimum bounding rectangle. The degree of texture is the mean value of the Laplacian for texture, described above. The scattering matrix is the covariance matrix of pixel positions, equivalent to an ellipse fit.

2) Boundary segments include: chain code; length; contrast.

3) Vertices include: position; number of boundary segments.

4) Holes have contour length.

5) Line segments have:
   - distance from origin (rho);
   - orientation;

- length;
- positions of end points.

6) Topological relations have pointers among regions, boundary curves and vertices, together with subset/superset relations for holes.

7) In order to retrieve regions with similar color, a history of the tree of segmentations is maintained.

Calculation of properties of merged regions is described. Various secondary features can be computed easily from primary features. Only primary features are in the data base. There are three functions for retrieving regions:

$$ALL-FETCH,$$
$$THERE-IS,$$
$$T-FETCH.$$

corresponding to all regions from a set with specified properties, the first region of a set with specified properties, and all regions adjacent to a region. In two results of segmentation, there are 339 and 391 regions, occupying about 90 KB for data structures.

There are four object classes, sky, trees, buildings with subobjects windows, and roads with subobjects cars. A plan image is generated by taking regions with large areas, called keypatches. The system tentatively merges all small patches to adjacent keypatches by choosing the highest score based on similarity of color and compactness of merged region. No semantic information is used in the merge.

The plan formulated by the bottom-up process is a set of object labels for keypatches and estimates of their correctness. The top-down process examines these interpretations and analyzes small, detailed structures in the context of large patches which have already been interpreted. When the top-down process makes a significant decision, the bottom-up process is activated to re-evaluate the plan.

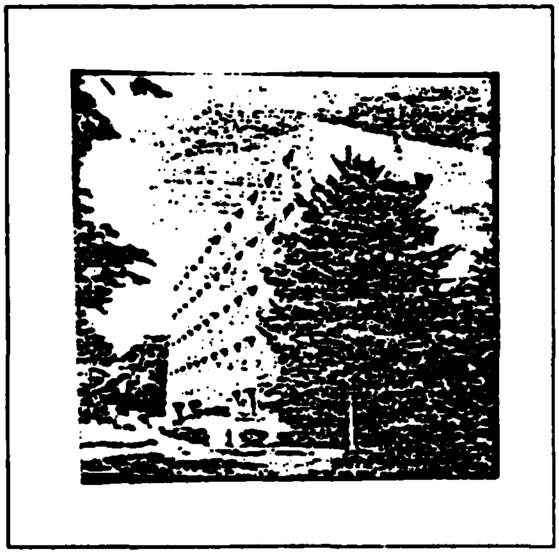
Rules for the plan are unary properties of objects and binary relations between objects. The plan manager computes a correctness value for every applicable rule applied to applicable regions in the plan image. Evaluation of relations takes place only after labels of regions are assigned.

In the top-down phase, each rule applied to each applicable region produces a score and an action to be registered on the agenda. At each step of analysis, the action with the highest score is executed and the database is changed. The agenda controls activation of production rules according to changes in the database. The agenda is updated whenever the database is changed. Each time, the number of tests is (the number of regions) times (the number of surfaces), i.e. several hundreds times several tens, a total of thousands. In order to decrease computation, a coarse-to-fine analysis is made, in a scene phase and an object phase. When a keypatch is labeled, the agenda activates the scene phase to re-examine keypatches which have not been interpreted. When a patch of an object is labeled, the agenda activates the object phase for that object to examine patches touching the patch just labeled. As a result, the number of tests at each stage is several tens.

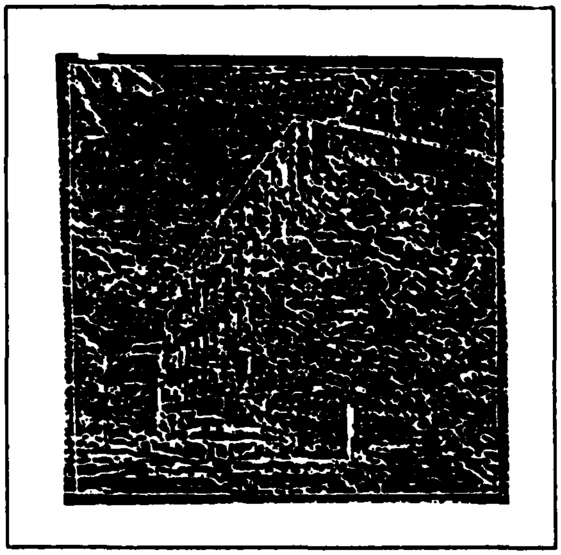
Each rule has a condition and an action. The condition is a fuzzy predicate. There are TO-DO and IF-DONE rules, corresponding to antecedent and consequent theorems of PLANNER [Hewitt 68]. Since only one region at a time is examined, there is a problem with global shape involving multiple regions. The system uses three mechanisms: the plan image; sets of patches retrieved from the data base; and special rules, e.g. extracting the shape of a building.

The world model is a network of knowledge blocks which define objects, materials, and concepts. Production rules are divided into subsets stored in particular knowledge blocks; the subset for scene level analysis is stored in the block SCENE, the subset to analyze objects is stored with the object.

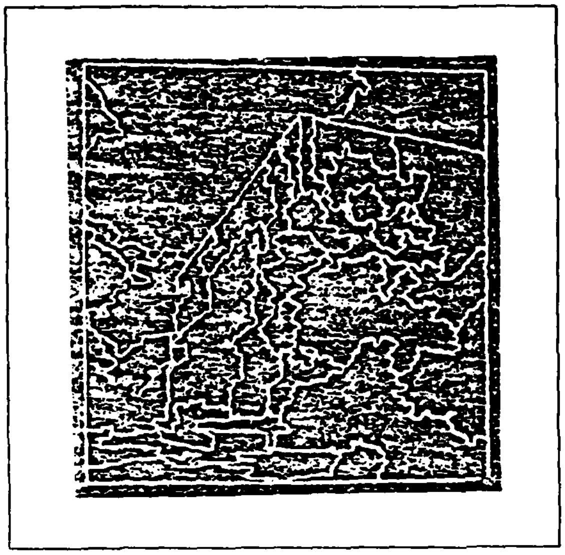
Results are shown for several scenes. Patches with area greater than 300 pixels are keypatches. There were 57 rules in total. Figures 2-25 through 2-48 show examples of processing three scenes. A region of the building is initially assigned a high correctness value for SKY because it is bright and grey. In the revised plan, it has a high correctness value for BUILDING because of the relation between building and sky. Horizon is detected by a production rule. The horizon appears to be found as the lower bound of the sky, by distinguishing sky from road.



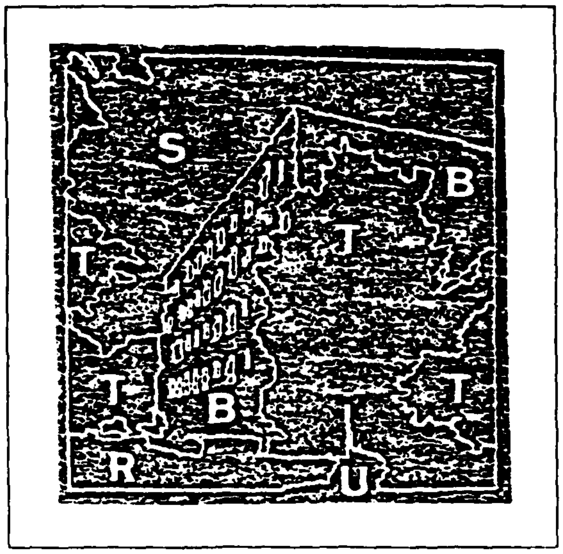
Digitized input scene
Figure 2-25



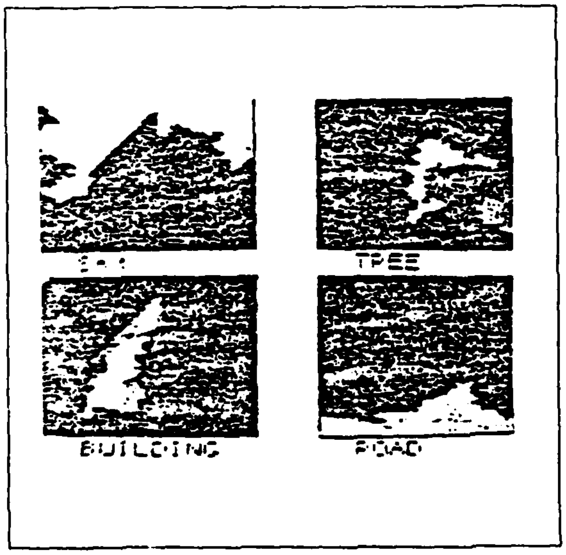
Result of preliminary segmentation
Figure 2-26



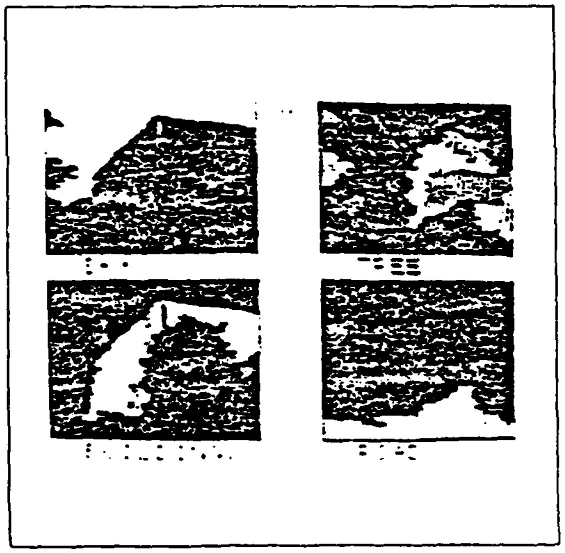
Plan image
Figure 2-27



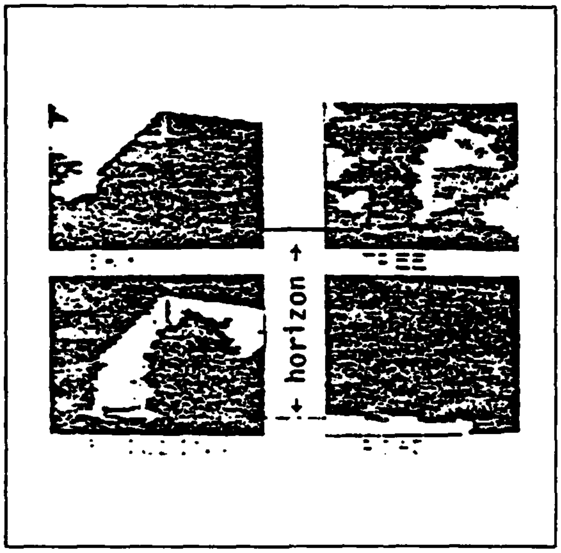
Result of meaningful segmentation
S: sky, T: tree, R: road, B: building, U: unknown
Figure 2-28



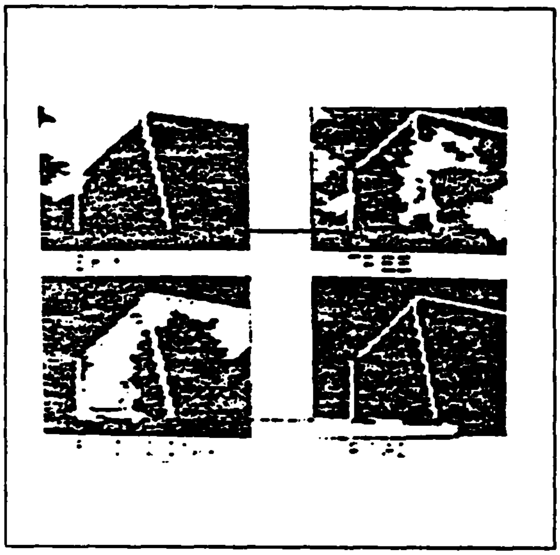
First plan obtained by using only the property rules
Figure 2-29



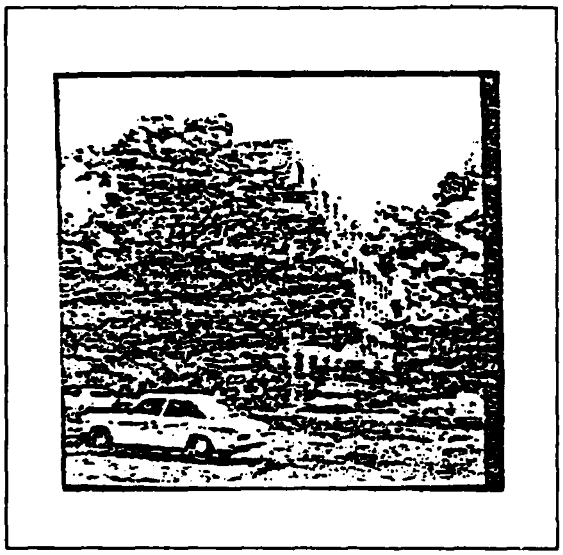
After using the relation rules
Figure 2-30



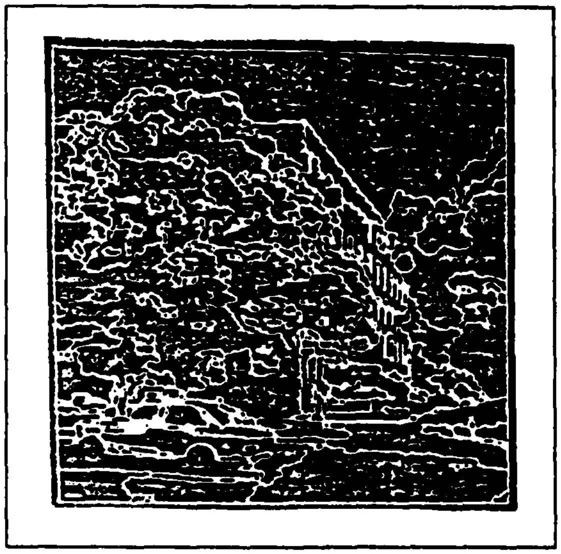
After extracting the horizon by the top-down analysis
Figure 2-31



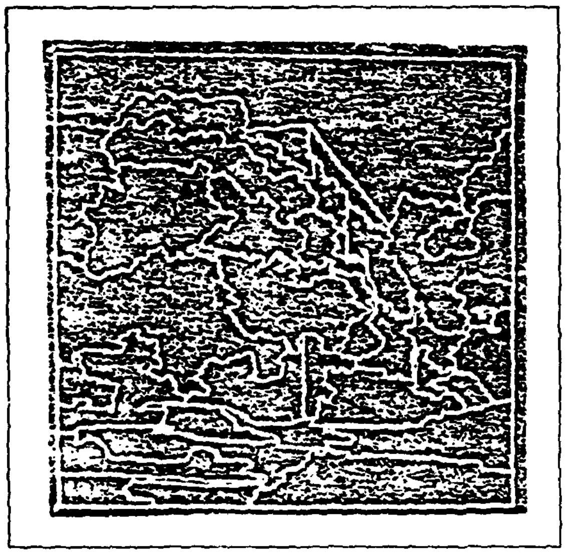
Outlines of the building by the top-down analysis
Figure 2-32



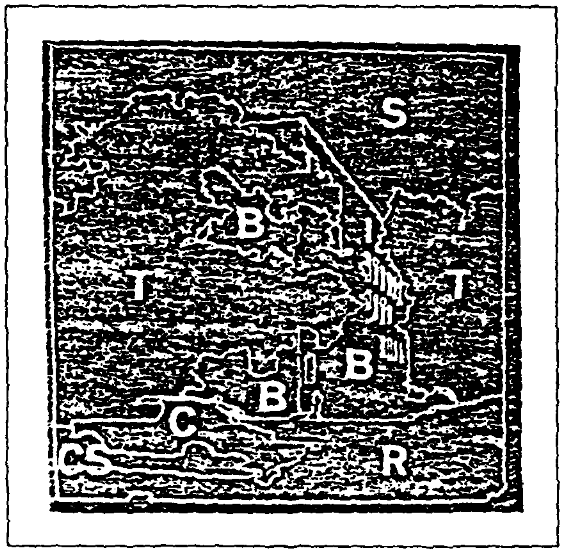
Digitized input scene
Figure 2-33



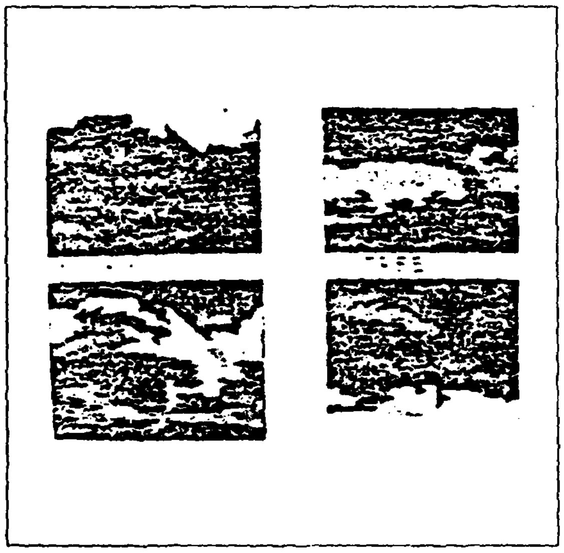
Result of preliminary segmentation
Figure 2-34



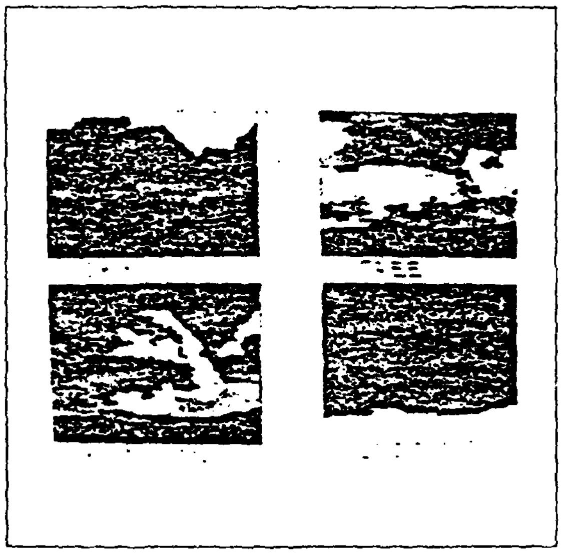
Plan image
Figure 2-35



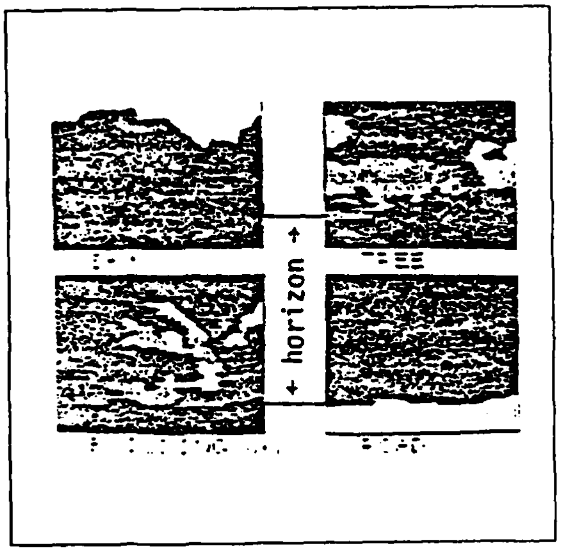
Result of meaningful segmentation
S: sky, T: tree, B: building,
R: road, C: car, CS: car shadow
Figure 2-36



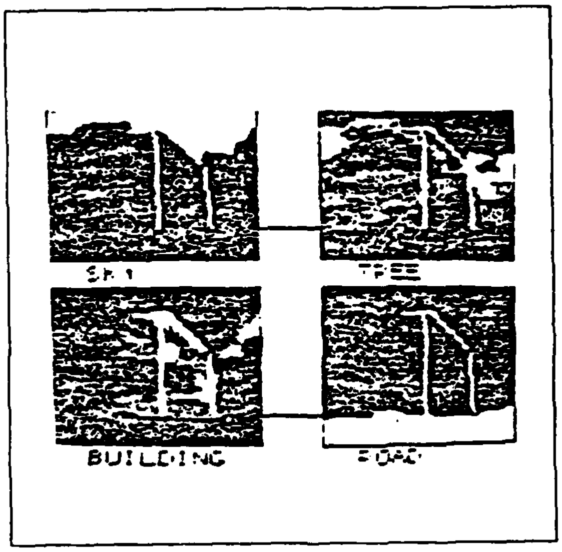
Plans generated for the scene. First plan obtained by using only the property rules
Figure 2-37



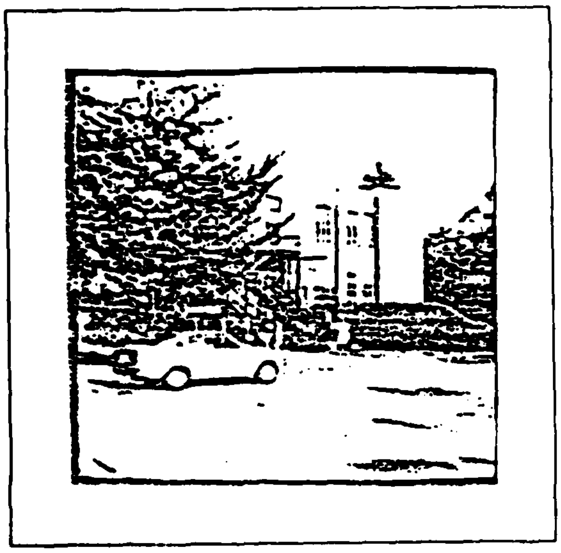
After using the relation rules
Figure 2-38



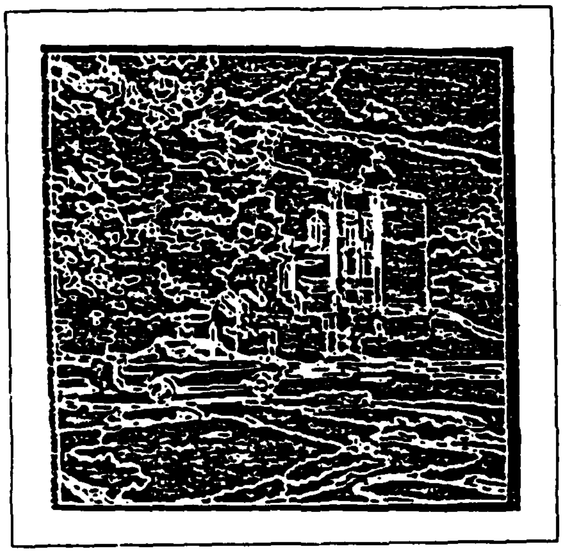
After extracting the horizon by the top-down analysis
Figure 2-39



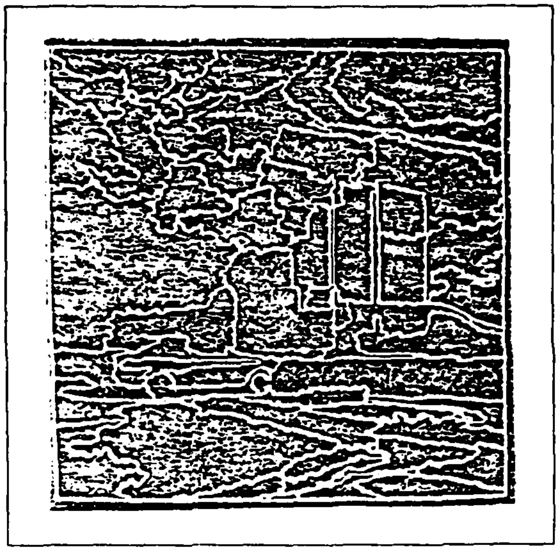
Outlines of the building by the top-down analysis
Figure 2-40



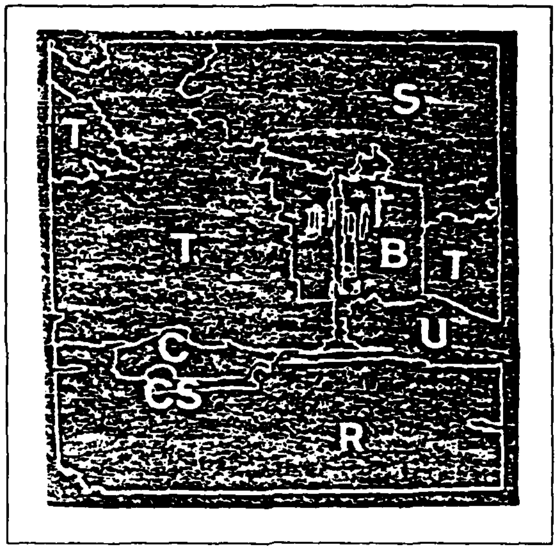
Digitized input scene
Figure 2-41



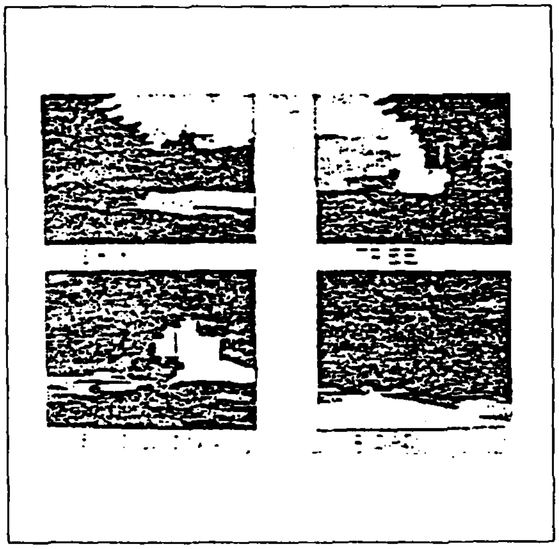
Result of preliminary segmentation
Figure 2-42



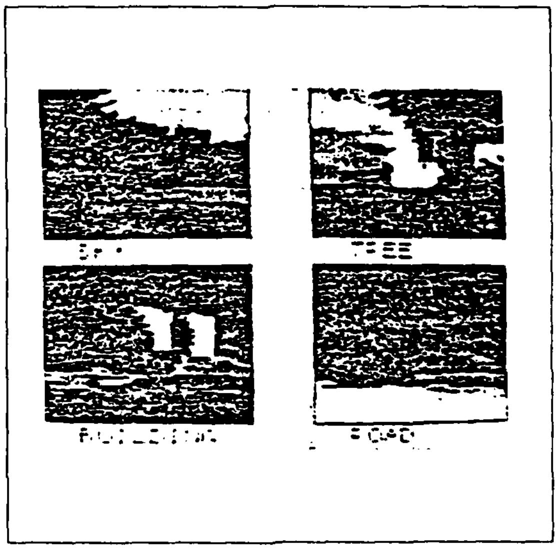
Plan image
Figure 2-43



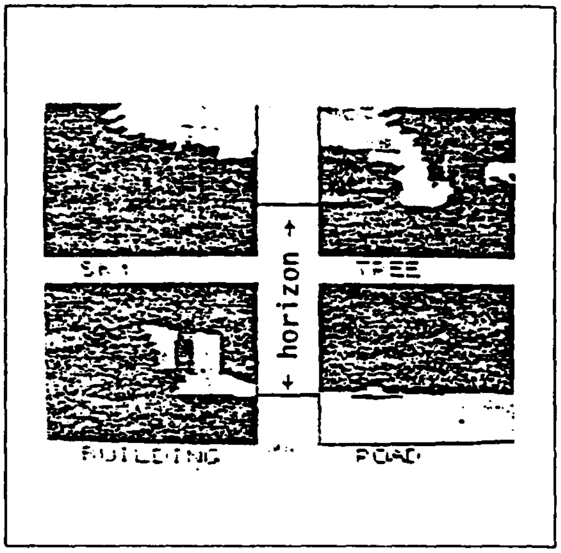
Result of meaningful segmentation
S: sky, T: tree, B: building, R: road, CS: car shadow, U: unknown, C: car
Figure 2-44



Plans generated for the scene. First plan obtained by using only the property rules
Figure 2-45

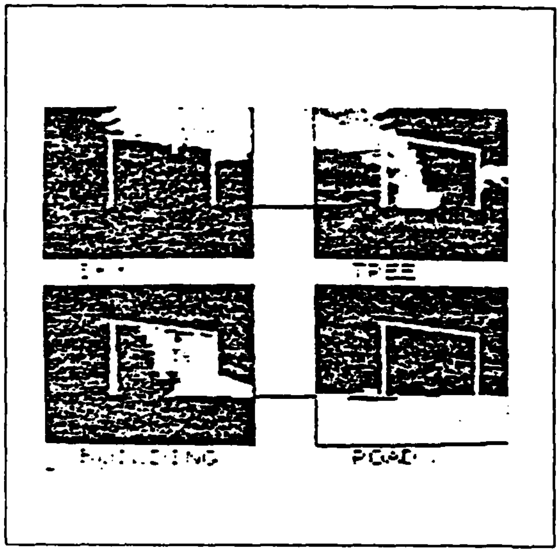


After using the relation rules
Figure 2-46



After extracting the horizon by the top-down analysis

Figure 2-47



Outlines of the building by the top-down analysis

Figure 2-48

The system does well overall. It demonstrates one of a few examples of reasonable performance on scenes of moderate complexity for a handful of images which are rather different. The thesis does not describe the rules themselves except by a listing, or help very much in analyzing the system performance and expected limitations. There are only a few objects in the model.

Models for analysis follow. The ROAD model has sub-objects: car, shadow. It is made of: asphalt, concrete. It has properties: horizontally long; touching lower edge of picture, and has relations: below horizon. The CAR in the road model is horizontally long; dark; and above the ROAD.

The SKY model has properties: not touching lower edge; shining; blue or grey; not texture; and touching upper edge. It has relations: linear boundary on the lower side.

TREE is made of leaves. It has properties: in the middle of the picture; heavy-texture.

The BUILDING is made of concrete, tile, or brick. It has sub-objects: WINDOW. Its properties are: in the middle of the picture; many holes; many straight lines; hole straight lines. It has a relation of linear upper boundary with the sky.

Some limitations of this system follow.

1. **The quality of segmentation is weak.**

1.a Thin linear features which do not show up in histograms are important, e.g. in identification of cars.

1.b The description of texture is weak.

1.c The organization of relations among patches is weak. For example, colinearity of window boundaries is not determined.

2. **Interpretation is image dependent.**

2.a Models are image dependent. The model for road (touching the bottom edge of the picture) is too specific to be useful.

2.b Models are weak. The model for car depends on a car being on the road. It has the relation that the car appears dark relative to the road, which is not adequate even for this domain. It makes weak use of shape. A human would identify the car make in many cases; the system has performance which is much inferior.

2.c There are models for few objects.

2.d The assumption that a single region does not include more than one object is not realistic.

2.e The approach is ineffective in many situations in which fine details determine object labels.

2.2.4 – ACRONYM; Brooks 1981

ACRONYM is an implemented interpretation system containing a substantial core of fundamental mechanisms which are powerful and general. Its performance demonstrated thus far depends on domain-independent capabilities, not on special domain-dependent tricks. ACRONYM as it stands is a large system which is part of a larger scheme for a general vision system. The author of this survey is biased by his enthusiasm for ACRONYM. ACRONYM aims to be a general vision system. This objective requires an enormous effort across all levels of a vision system, an effort beyond the state of the art. Substantial progress has been achieved. Immediate plans call for mechanisms which will greatly increase its power and breadth.

ACRONYM's objectives include:

1) three-dimensional interpretation;

2) a rigorous scientific basis;

3) high performance: important applications are typically difficult; care in building high performance modules enables generality;

4) general: it is intended to provide an option for a standard system with a user base providing technology transfer; 4.1) a large general core of powerful capabilities cn which to base applications;

4.2) interpretation generic with respect to object class, providing commonality of programs for similar applications, and providing a means for inspection distinguishing essential vs cosmetic flaws;

4.3) interpretation generic with respect to observation, i.e. insensitive to viewpoint and flexible with varied sensor inputs;

4.4) mechanisms to use special case information and data.

5) Complete: interpretation making use of total information, data, and knowledge; including multi-sensor data, knowledge of experts; geometric models of objects, prediction of

expectations; incorporating powerful.use of shape in modules for edges, texture and image organization, depth, shape from shading, and shape from shape; incorporating mathematics and physics.

6) System: a critical mass effort over a continuing period, seeking collaboration and applications.

7) Engineered for easy use: input in the form of geometric models and rule bases; automatic programming; user aids.

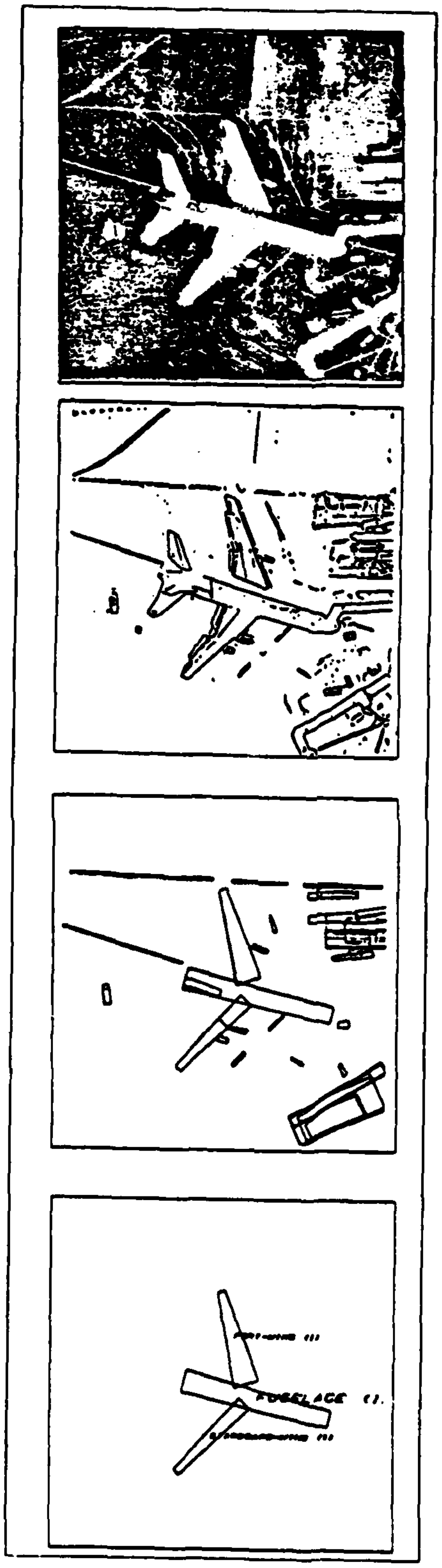
A user should program ACRONYM in terms of geometric models and geometric task specifications, a common language natural to both user and ACRONYM. The user refers to models in a geometric data base and shows examples of typical members of the object class. ACRONYM should infer specific and generic properties of the object class and contexts, especially causal relations. ACRONYM constructs perceptual programs to carry out the assigned task. ACRONYM should be general and generalizable in the sense that problem-specific information can be embedded in general problem-independent mechanisms which provide a natural decomposition of problems into physically meaningful elements. A core built up from a few problem domains should cover most capabilities required for other domains.

ACRONYM has been demonstrated successful on a few images of aircraft in aerial images, not enough to warrant a claim of generality. Figure 2-49 shows an example with high resolution. Figure 2-50 shows an example with poor resolution. The ribbon finder has problems with the poor resolution image, but ACRONYM recognizes the three aircraft for which ribbons are reasonable. In another image like Figure 2-50, the ribbon finder did not turn up any adequate ribbon descriptions.

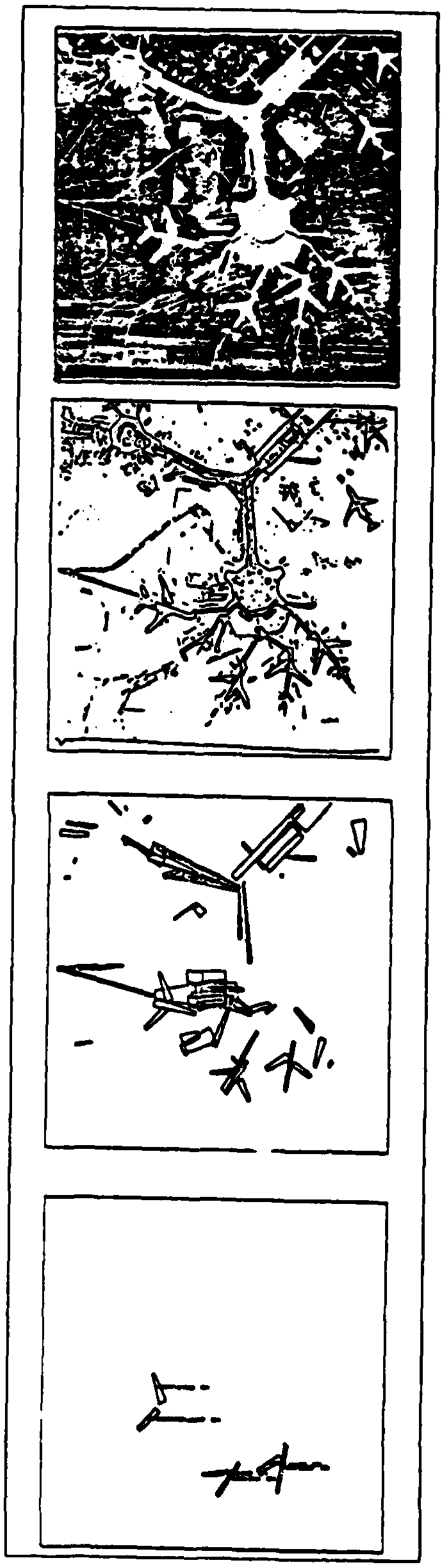
In a real sense, ACRONYM reasons from first principles. ACRONYM has a general core in that its rules implement algebra and projective geometry. There are no special rules for aerial images or aircraft. There is no profound reason why ACRONYM could not recognize aircraft in images taken at ground level although it will probably break when tested on such images, because of bugs or missing capabilities which were not exercised previously. For example, at ground level, the fuselage appears more or less the same as from above; wings are less observable, but engine pods and tail are more prominent. The rule base will need to be extended considerably in dealing with varied object classes, e.g. manufactured parts, vehicles, and buildings.

ACRONYM has viewpoint-independent three-dimensional object models in the form of part/whole graphs, in terms of generalized cylinder primitives. ACRONYM represents object classes, for which subclasses and specific objects are represented as restrictions by constraints in the form of symbolic expressions with numeric type.

ACRONYM searches for instances of models in images. It employs geometric reasoning in the form of a rule-based problem-solving system. Geometry is the key, while the rule-based system is simply a way of implementing geometry. We think that a formal representation of geometry is necessary to make a compact and coherent set of rules, to get additivity and consistency in a rule base. Despite claims to the contrary, it seems clear that a rule-based system of itself does not aid in making additivity and consistency of reasoning. We believe that building a vision system is 1% a system effort of the sort which are familiar in computer science, and 99% basic science.



ACRONYM example 1
Figure 2-49



ACRONYM example 2
Figure 2-50

ACRONYM predicts appearances of models in terms of ribbons and ellipses. It uses an edge finder to make observations of ribbons and ellipses in images. ACRONYM finds observed ribbons consistent with predicted ribbons, and restricts interpretations to those which are parts of clusters consistent with predicted structures of ribbons. It interprets in three dimensions by enforcing constraints of the three dimensional model. Thus, to identify aircraft, it matches observed ribbons to predicted ribbons for wings and fuselage, then finds clusters of ribbons which are consistent with combined wings and fuselage of a three-dimensional aircraft model.

ACRONYM makes predictions which are viewpoint-insensitive, in the form of symbolic constraint expressions with variables. One mechanism of viewpoint-insensitive prediction is the use of observables which are invariant and quasi-invariant over large ranges of viewing angle. ACRONYM does not generate all possible views of an object. Total prediction has combinatorial complexity. For a polyhedron with n distinct faces, there are 2^n views. Instead, ACRONYM predicts partial views for individual faces, of which there are of order n. Coherence of several features comes from merging constraints. The image shape descriptors are invariant under image rotation. To generate predictions, ACRONYM starts with models in their coordinate frames. All contours of faces which might be visible are identified. They are transformed into the camera coordinate frame with symbolic translations and rotations represented by variables. The system simplifies the corresponding expressions, and makes a projective transformation. Those contours which are visible are identified. Relations between these contours are predicted, and the shape of the contour is predicted. Back constraints which relate image observables to model parameters are generated. Relations between ribbons are generated. Each ribbon provides a number of back constraints which combine to constrain model parameters. ACRONYM automatically determines ribbons and ellipses for parts of the object which it determines are most observable. Predictions of feature shapes are nodes of the prediction graph; arcs of the graph are image relations between features. Arcs relate multiple feature shapes predicted for a single cone. The swept surface and end surface of a cone are predicted separately. A prediction is made that they are coincident.

Interpretation proceeds by combining local matches of ribbons into clusters. Global interpretations must be consistent in two ways: First, they must satisfy constraints specified by the arcs of the prediction graph. Second, accumulated constraints that each local match imposes on the three dimensional model must be satisfiable. The interpreter searches for maximal subgraphs of the observation graph which are consistent with constraints of subgraphs of the prediction graph. Each such match is an interpretation graph. The interpreter matches ribbons against predictions of ribbons. It then tries to instantiate arcs of the prediction graph by checking pairs of regions to see whether they satisfy relations predicted. For pairs which satisfy image relations, it merges constraints on the underlying three-dimensional model.

ACRONYM has been used as the basis for a simulator for robot systems and for automated grasping of objects, with a rule base for determining which surfaces are accessible in the initial position, which surfaces are accessible in the final position, and ways to grasp with maximal stability.

The top-down paradigm is only one part of the ACRONYM design. A top-down system is far from general. It is believed that this paradigm can provide only a small part of human performance, even though prediction has been made relatively powerful in ACRONYM by use of predictions which are generic with respect to object class and viewpoint-insensitive. We believe that the way to a general vision system lies in spatial understanding, as contrasted with image understanding. That is, prediction of images and matching at the level of images is inherently limited, a convenient expedient reflecting the weakness of our descriptive mechanisms, but it is not a fundamental approach. Instead, we believe that the major part of interpretation is not at the image level, but at the level of volumes. Descriptive mechanisms which generate volume descriptions are essential, combined with prediction and interpretation at the level of volumes. Certainly stereo, motion parallax, and object motion are important observation capabilities, together with shading. Recent theoretical work on monocular interpretation of surfaces from images [Binford 81, Lowe 81] make it appear promising that general mechanisms for generating spatial observations from images will be developed soon to support general vision systems.

Limitations of ACRONYM as a general system follow.

1. **ACRONYM has weak segmentation.**

1.a The ribbon finder determines spurious ribbons. It misses small ribbons.

1.b Grouping of ribbons is not done in segmentation, only in interpretation.

1.c The linefinder and ribbon finder perform badly with texture.

2. **Interpretation is limited.**

2.a Image prediction and matching is not sufficiently general for scenes with many objects.

2.b It has been tried with only a few objects and has models for very few.

2.c It tests all pairs of ribbons in establishing relations, ignoring proximity.

2.d It has been tried on only a single viewpoint.

2.e The top-down paradigm is inadequate for complex scenes.

2.2.5 – Shirai 1978

Obvious edges are found in the entire scene; they are described by straight lines or ellipses. Edge points are found by using one-dimensional profiles and classified into three types. Averaging is done over small areas, typically 3x3. The direction of the edge is determined from the gradient. An edge kernel is determined by a set of edge points of the same type with similar gradient directions. Continuations of edge kernels must have the same edge type. The tracking phase predicts an edge element and verifies it. Tracking may insert a fictitious edge point at the predicted position. Tracking proceeds in both directions until both ends are terminated by connecting to another end. Some edges are extended to fill small gaps.

Curve description has two phases, segmentation and curve fitting. Segmentation uses curvature vs arc length along the curve, where curvature is defined as the angle difference at a center point between chords a fixed distance on either side (approximately the difference of tangents). The routine finds sequences of high curvature to place knots. It tries to classify curves: if sagitta is large, it is a curve; if angle change is small, it is a line. It tries to merge adjacent undefined or curved segments. A method symmetric in x and y is used to fit curves. Curves are fit with ellipses. If the search does not converge, the number of parameters is decreased, i.e. successively fitting a circle and straight line.

Analysis of the scene starts from the most obvious object, then the next. For object recognition, find the most obvious feature, then find a secondary feature to verify identity and determine object range. Then find other lines on the object. For recognition of a lamp, the program locates the lamp shade, then looks for the trunk of the lamp which supports the shade, then the lamp base.

The objects include a lamp, a book stand, a cup, a telephone, and small objects (pipe, pen, etc). For the lamp, the primary feature is a bright elongated strip for the lamp shade of a fluorescent. Secondary features are a pair of vertical edges corresponding to the trunk, and the contour of the lamp base. The primary feature for the book stand is a cluster of long vertical lines in a rectangular region. Secondary features are lines connected to the verticals. The cup has a pair of vertical edges; secondary features include the ends of the cup connected to the verticals. The telephone has an ellipse for the dial and one outside the numerals; secondary features include features surrounding the ellipses. Small objects have shape and size of contour as primary features; secondary features are shape details and light intensity surface.

Small objects such as pens or erasers are tried after large objects are found or many edges are obtained. Otherwise a small object might be confused with part of a large object. The system finds more edges by decreasing the reliability level and looking in delimited areas. The system may take a close-up image where necessary.

Experiments have been conducted successfully with a range of positions and orientations of objects and varied lighting conditions. Figures 2-51 through 2-58 show an example from [Shirai 78]. As with previous systems, there are substantial limitations of this system as a general vision system.

1.a The edge finder is adequate for this task, but it would not perform well in complex scenes.

1.b The system cannot deal effectively with texture.

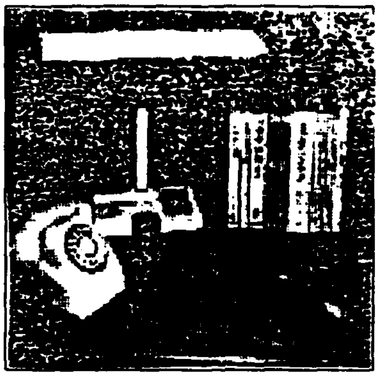
1.c The system has no organization of related edges.

2. Interpretation is likewise limited.

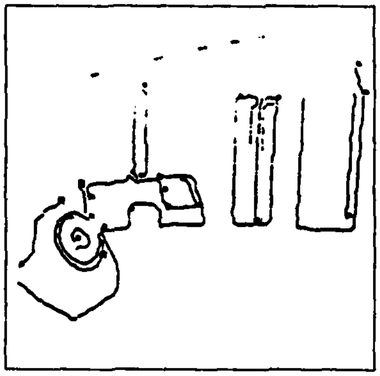
2.a There are only image models, not object models.

2.b There are only few objects.

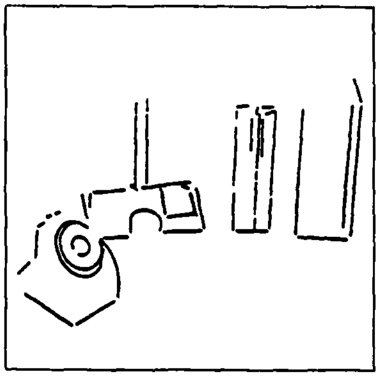
2.c The analysis is top-down, which is reasonable for few objects.



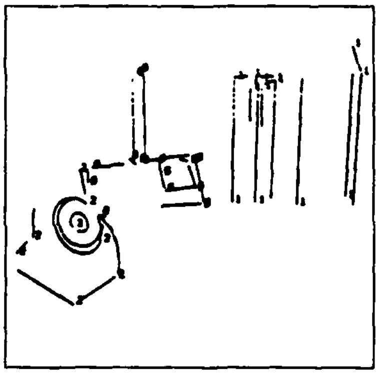
Original scene
Figure 2-51



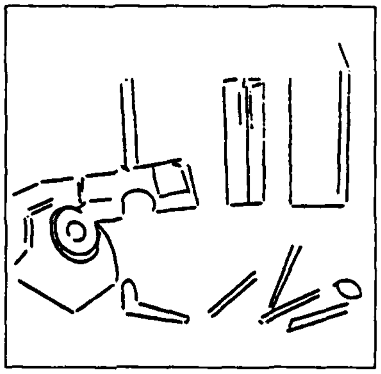
Edges found in cycle 1
Figure 2-52



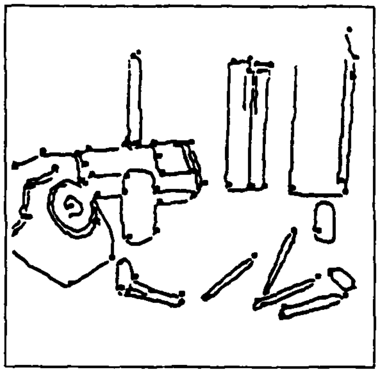
Description
Figure 2-53



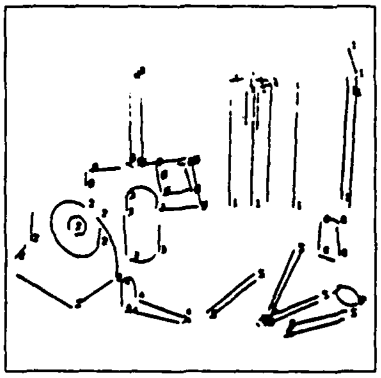
Recognition
Figure 2-54



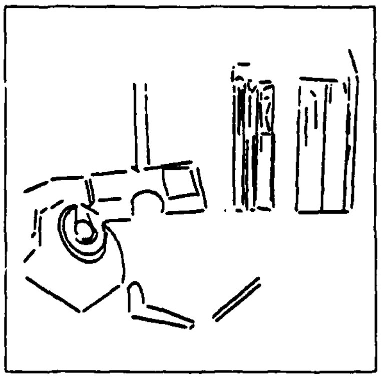
Description in cycle 2
Figure 2-55



Edges found in cycle 4
Figure 2-56



Recognition
0: lamp, 1: bookstand, 2: telephone, 3: cup, 4: pipe, 5: pen, pencil, ballpoint pen or felt pen, 6: ink pot, 7: eraser
Figure 2-57



Description of edges found without feedback process
Figure 2-58

2.2.6 – Ballard, Brown, and Feldman 1978

[Ballard 78] use image models in locating ships at docks and in locating ribs in chest x-rays. Geometric constraints are used. The system is oriented to answering queries; the level of detail is determined by the query. Only a portion of the image is interpreted. The system is structured in levels: 1) the model; 2) the sketchmap synthesized during image analysis; it relates the model and the image; 3) image data structures including images at different resolutions and spectral components, texture images, edge images, etc.

The structure is similar to VISIONS. Perhaps the main difference is that in VISIONS, segmentation is made to a level determined by the model, so that the image will be understood as fully as possible. Here, the query determines the level of detail.

Links encode constraint relations. They include the probability that the relationship holds, and include the expected value of the relationship. For example, SHIP ADJACENT DOCK is a constraint with probability and expected distance. A node may have a cost for evaluation depending on its operand nodes. This allows cost-benefit analysis. Locations are delimited by union and intersection. Constraints are two-dimensional.

Control involves synthesis of a sketchmap. Queries take the form of user-written executive programs. Procedure invocation is based on a description of a procedure's capabilities, together with preconditions and post-conditions. The executive decides which procedure to run based on cost-benefit, i.e. the lowest cost procedure which meets preconditions. Procedure descriptions include: slots which must be filled; slots the procedure can fill; cost and accuracy of the procedure; and a priori reliability of the procedure.

The user program is responsible for 'strategic' resource allocation beyond the executive level. No single domain-independent problem-solving is used.

One example is locating docked ships. A photo with matching map is used; registration is done manually. The constraint is used that ships are parallel to docks at half the ship's width. Template matching was the only visual technique; the center is estimated midway between the locations at which correlation goes above threshold and goes below threshold.

An example is shown of finding ribs in chest x-rays. Only lower edges of ribs are found. There are procedures to locate parabolic rib segments, to translate and verify adjacent ribs, and to translate without verifying.

The representations used are:

1) straight lines: ordered list of points; list of segments; circular lists of points;

2) region boundaries: y list consisting of a y value followed by x values for entering and exiting region.

The operators included are: correlation template matching, Hough, Hueckel, distance of point from segment; segment parallel to segment; union and intersection of regions.

The system is special case and not intended to be a general vision system. Its limitations as a general vision system are many.

2.2.7 – The Verification Vision System; Bolles 1976

The Verification Vision (VV) system [Bolles 76] uses object models and image models. It is intended for inspection and visual control in repetitive manufacturing tasks. VV makes use of three-dimensional models, but it requires sensed images to be very similar to image models. Thus, for a camera which is one meter from an object, the position of the object may vary by about plus or minus one cm, and its orientation may vary by plus or minus fifteen degrees, but there may not be major shifts in appearance or major changes in occlusion. The major reason for that restriction is that VV's primary visual operator is area correlation of small windows of an image with reference windows [Moravec 77].

An object model contains a set of point features and their locations in three space. The image model includes a set of training images and features in those images. Features are small windows centered around elements determined by an interest operator [Moravec 77]. This is a weak object model. From the training set the system builds an estimate of the variations of locations of features and an estimate of the effectiveness of its feature detection operators on the designated image features.

Four stages are distinguished:

1) Programming time: the user states the goal of the task, calibrates the camera, and chooses potential operator/feature pairs.

2) Training time: the system applies the operators to several sample pictures and gathers statistical information about their effectiveness.

3) Planning time: the system ranks operators according to their expected contribution, determines the expected number of operators needed, and predicts the cost of accomplishing the task.

4) Execution time: the system applies operators in a order of their cost-effectiveness, combines the results into confidences and precision, and stops when desired confidence has been achieved or cost limit exceeded.

The goal of the task includes achieving sufficient confidence of correct correspondence, adequate precision of location, and sufficiently low cost of achieving required confidence and precision. VV uses least-squares fitting to combine results of multiple measurements to estimate position and precision. Since least-squares fitting is sensitive to assignment errors, a Bayesian estimation scheme is used to choose assignments with few errors. Two sorts of information are used in making correct matches of image features to reference features: 1) confidences obtained from values returned by operators; 2) relative image position of subsets of features. Because of combinatorics, the second method must be used with small subsets of features. Later, a search for maximal cliques consistent in three-space separations was used. The latter approach is much more satisfying conceptually, since relative image distances are not invariant and their errors are not predictable without use of a three-dimensional model.

Two types of features were studied:

1) corner-like features obtained by an interest operator and matched by a binary-search correlator [Moravec 77];

2) curve features predicted by a geometric model of the object [Miyamoto 75] and determined by a curve verification procedure based on the Hueckel edge operator [Hueckel 73]. The latter ware not integrated into the VV system.

The programming of VV problems was automated considerably. The program produces a set of features. The programmer filters out suggested features that are judged unreliable. In gathering statistics, VV displays a reference picture side-by-side with a training picture, with matching features marked. The system makes its best assignments according to consistent subsets. It asks the user for confirmation. It ranks operators according to an ad hoc quality measure. In execution, it applies operators in order of the measure.

The system was limited in that it primarily depended on small correlation windows as features. Thus, it was restricted in viewpoint. It was intended for a few objects. It used three-dimensional models of objects, but they were point models, without shape relations.

2.2.8 – Faugeras and Price 1980

In [Faugeras 80], the input is a network of segments from procedures for region-based image segmentation and linear feature extraction. There are about 100-200 total segments of both types. Properties of segments include average color, simple texture measures, position, orientation, and simple shape measures. Relations between image segments include adjacency, proximity, and relative position. The model description is identical to the network of image segments. If a relation occurs in the model, it is expected to occur in the image description. The model is not a complete description of the scene, but apparently an image model.

Matching is a graph endomorphism. Solution is by stochastic matching (relaxation). Initial compatibilities are computed without relations, because labels are not known; they combine weighted differences, taking up to 30 possible labels, or those with compatibility greater than 1/10th of the best, whichever is fewer. The compatibility measure with relations is computed only with the most likely assignments for the second object (usually only one assignment). Compatibilities are computed as needed. The compatibility measure is the dot product $p \cdot q$ where p is the probability vector $(p(l_1), p(l_2)..)$ and q is the prediction vector $(q(l_1), q(l_2), ..)$ for labels l_i.

It becomes a constrained optimization approached by steepest descent. Macro-iterations of steps are composed in order to make decisions to assign names to units with high probability (above 80%).

It is hard to tell what its performance is for the two images shown. The images are not shown in this survey because they do not show very much.

The system has image-dependent models, and is strongly restricted in viewpoint. Segmentation is relatively weak.

2.2.9 – Garvey 1976

In [Garvey 76], the system's function is to locate objects in an office environment, in which it is assumed that all objects are known. The system has as input measurements of range, reflectivity at one wavelength, and three color images. The system's strategy is to acquire image samples which might belong to the object, to validate the hypothesis, and to bound the image of the object. The system uses simple, local features rather than structured shape descriptions. It uses contextual relations, e.g. a telephone is on a desk, to decrease search and to minimize the set of possibly ambiguous objects. The approach follows a belief that in a sufficiently restricted environment, a set of local distinguishing features can be found which are effective in initial screening for candidate matches. This rests upon having models for all objects and not having many objects.

The system searches for cost-effective strategies of sequences of operators. In acquisition, it chooses an appropriate limited search window of the image, sampling the image at a density determined by the object's size, maximum range, and least favorable orientation. Estimates are made of the cost of search with various operators, together with likelihood that the operator will be successful and will be correct. Which operators are effective depends on all objects in the image.

The system is programmed interactively. Objects are shown to the system by outlining them in an image. Objects are automatically characterized by conjunctions of histograms of local surface attributes such as hue, orientation, range and height, and relationships between surfaces. These characterizations provide ingredients for strategies for object finding.

In experiments, its performance rests strongly on having depth data and surface orientation derived from depth. For finding a desk, the desktop has a height of 2.5 feet and horizontal orientation. For a door, height, orientation and hue were not enough. Its size and location, together with a vertical rectangle characterization were essential.

Regions are represented as lists of samples, as a list of vertices of a closed polygon boundary, by a bounding rectangle, and by vertical and horizontal bounding rectangles in space. Objects are represented by distributions of attributes of surfaces, with shape of surfaces and relations between surfaces.

Some limitations of the system follow from the approach of distinguishing features. Choosing simple features which distinguish regions corresponding to objects of interest assumes that there are few objects and all are known. The interpretation is strongly dependent on depth data and probably would not achieve similar performance without depth data. Shape, other than that obtained from depth, is used in a weak way. Also, the top-down approach does not go far in a world of many objects.

2.2.10 – Levine 1978

[Levine 78] describes a three-level system, of which the first two levels were implemented. The first level segments pictures into regions without scene context. The second level has a local phase and a global phase. The local phase matches image templates with observed image regions, using A^* graph search. It performs global optimization using dynamic programming to merge regions and assign labels to them. The highest level design includes a standard relational data base and a system like production rules.

The system is implemented using MIPS, an interactive image processing system. Two data types are used uniformly throughout: image arrays and feature vectors. Low level region analysis obtains about 200 regions from a 256x256 image, a reduction of about 300:1 in items, not necessarily in storage.

The intermediate level deals with standard views. The output of the intermediate level is an ordered list of interpretations for each region. The local intermediate level process is a form of template matching of object descriptions. It is model-driven and uses heuristic graph search to match shape, color, and texture. The second process is global, also model-driven. It incorporates spatial relations as a global optimization problem solved by dynamic programming. Model input for this level comes from interactive designation of regions and objects, computing features, and updating the data base. Symbolic information in the form of text is also a form of input.

The high level system design includes a vision production system with relational database.

Low level segmentation is based on regions found by a shared nearest neighbor clustering modified by connectivity. Processing is approximately in order of decreasing size using a pyramidal data structure. The edge pyramid has edges from a gradient operator. The pyramid thickens edges until at the top level, all regions have been extinguished. Areas of the picture which are farthest from edges tend to be extinguished high in the pyramid. From a starting region, projecting down to the next level involves expanding a pixel into four pixels. Those which are marked as edges are not examined at this level, but are looked at in the next level down. At this level, those not marked as edges are marked for clustering using the nearest neighbor algorithm.

Local template matching is used to match collections of adjacent regions against all stored object prototypes. Features are stored in three classes according to decreasing importance in reducing search time. The first class includes the minimum bounding rectangle, and its area. 'Intrinsic' features make up the second category: intensity, hue, saturation, and texture. The third category includes six moment invariants as a rough measure of shape, and detailed shape from a set of Fourier coefficients for the outline, used only in final template evaluation.

The A* algorithm is used for graph search with an evaluation function. Nodes are regions, while arcs are region adjacency. The estimate of the cost to node n is the number of nodes expanded from the start.

Relational information is used in the optimal search stage, done with dynamic programming. Regions may be ordered either according to region area or decreasing max confidence value among interpretations. The orderings of model and data do not necessarily correspond; no mention is made of whether the algorithm accounts for the difference in order. The transition function is a linear weighting of differences between structural relations in the image model and observed relations. Relations include $LEFT-OF, RIGHT-OF, ABOVE, BELOW, ADJACENT-TO, CONTAINS$, and $CONTAINED-BY$.

The knowledge database or long term memory (LTM) in the high level system design is a management-type relational database with accessing operations which constitute a relational algebra. Operations such as JOIN, PROJECT, INTERSECT, UNION, and RESTRICT are available. The high level system will be a data-driven production system. Thus far, it only deals with images, not three dimensional scenes. The LTM has subworlds for outdoor scenes, office scenes, etc. Objects in LTM have associated actions to be taken under conditions of the STM. A short term memory (STM) contains a list of regions and their interpretations. It resembles the 'blackboard' of HEARSAY. Regions are interpreted sequentially, unless an action is involved which alters the sequence. For each region, there is a list ordered by decreasing confidence. Implicit actions are invoked by the system when a region matches an object in the LTM with a confidence above threshold. Explicit actions are invoked if there is only one interpretation for a region.

Limitations of this system are found throughout. Some come from the segmentation process which relies on a gradient operator, with all its weaknesses. The intermediate level is viewpoint-dependent. Apparently, the top level will be built upon the intermediate level, and will also be viewpoint-dependent.

2.2.11 – Parma, Hanson, and Riseman 1980

[Parma 80] take a color image of an outdoor scene and an image model with three dimensional relations and build a symbolic model of the three-dimensional world shown in the image, in the form of names of objects and weak relations in three-space.

The model contains image locations of objects, uncertainty radii, and size radii. They emphasize representation of knowledge structures and control of knowledge sources (KSs). Their knowledge structures are schema for a scene concept, e.g. road scene or house scene, with control for invoking a subset of knowledge sources. They believe that schemas provide a bridge between general purpose and special purpose systems. They raise issues of control:

1) basis for invoking KSs;

2) using alternative hypotheses provided by KSs. The basis for invoking KSs is top-down control driven by schema.

They identify a set of experiments with varying generality:

a) Specific scene schema from a known viewpoint; specific 2-D schema can be obtained; the example of this report was a particular house from a known viewpoint;

b) General scene schema, known viewpoint; I don't understand what they mean; how can we know viewpoint for non-specific scenes?

c) Specific scene schema, unknown viewpoint;

d) General scene schema known, unknown viewpoint;

e) General scene schema unknown;

f) No scene schema; construct a partial 3-D surface/volume description; this is a subset of the general vision problem, since it does not use familiarity information.

Two forms of segmentation were described: edge relaxation; and region relaxation using histograms. Both were implemented in a simulation of 'processing cones'. Both use two complementary relaxation labelling processes: boundary formation: local differences; region formation: global similarities. Edge relaxation finds local discontinuities in a feature (intensity or color) along horizontal and vertical edges between pixels. Iterative relaxation on small neighborhoods forms boundary segments. Region analysis proceeds by cluster detection of peaks in the histogram of one feature or in the joint density function of pairs of features. Connected sets of pixels are determined as regions.

Eleven modular knowledge sources (KSs) were included, implemented in a graph processing language built in ALISP. They include: Inference Net KS; 2-D curve fitting KS; 2-D Shape KS; Occlusion KS; Spectral Attribute Matcher KS; 3-D Shape KS; Perspective KS; Horizon KS; Object Size KS.

They use a data base including a long-term memory of world knowledge which is not image-specific, organized into schemas, objects, volumes, surfaces, regions, line segments, and vertices (RSV regions, segments, vertices). Nodes are those levels, arcs are relations, primarily AND/OR. Inter-level arcs exist. Interpretation is viewed as a set of instantiations of nodes in LTM, put into STM. Short term memory is image-specific; it is used for constructing an interpretation Initially, it contains only the RSV levels.

2-D curve fitting uses splines with knots determined by places of high curvature measured as angle from a point to k-neighbors on either side. If straight line segments fit, they are used, else quadratic, else cubic. In the system, 2-D shape classification is hierarchical. For straight line segments, quadrilaterals include trapezoids and parallelograms which include rectangle and rhombus. Quadratics are used for ellipses and circles. Other types of curves are labelled as blobs.

Occlusion depends on continuous curve generalization of T junctions. It is not clear that it is used in the experiment described later.

Some non-manmade objects have relatively invariant color and texture. Many objects such as man-made objects vary in spectral characteristics. It will sometimes be right to guess among the class of 'target' objects, those with invariant spectra. In the example below, of 21 regions assigned on the basis of color, 11 were correct. Of the remaining, 5 were wall assigned as sky, and 2 were roof assigned grass.

They describe problems with inference nets. Probabilities are assigned to nodes and conditional probabilities to arcs, providing weighted paths by which implications of local hypotheses may be propagated up and down through the layered network. They point out problems in consistency and loops when generalizing.

The 3-D shape KS uses blending functions of cubic splines, defining 'quilted solids'. It is stated that 3-D shape has not been integrated into the system used in the experiment described in this paper. A specific 3-D schema can be transformed into a specific point of view and projected onto the corresponding image plane, with hidden lines removed. Those facilities were not available at the time of the experiment, but were available when the experiment was completed. Specific 2-D schema were used.

2-D Schema and 3-D schema have for each region a centroid of the expected location of its center and a radius representing decreased likelihood of the region center appearing at that location.

The perspective KS assumes a horizontal ground plane, with surfaces either vertical or horizontal. The camera model includes angle of inclination, image distance, and height above the ground. It deals with elevation, height, width and range. The assumption is sometimes made that an object stands on the ground. If the ground is planar, objects are on the ground, and if objects are identical, the horizon line can be determined. It is in the ground plane. Tilt is given by the angle of the horizon from the center of the image. If tilt of the observer were zero, the horizon would fall in the center of the image. Range of objects is determined by projected distance to their feet. If other objects are the same height as the camera, e.g. eyes of other people, a third plane is defined by a least rms fit; the plane goes through the horizon. It is not clear from the article whether those capabilities were implemented. Sky regions cannot be below the horizon line; grass regions cannot be above the horizon line. The ground is assumed planar, the horizon a level line in the Horizon KS.

Object size KS: If the perspective KS gives an estimate of object size in three space, the object size KS generates a list of object hypotheses ordered by confidence based on the region size. Perspective KS returns computed size and range of size; default range is 5%.

There are sometimes boundaries with known characteristics (e.g. long and straight, bounding roof). Top-down control of KSs directs matching of schema regions to image regions and some schema line segments to image line segments. A heuristic weighted evaluation function is left unspecified.

Experiment one matches color and texture attributes to improve a fragmented segmentation, by the rule of merging adjacent regions with the same object labels. A spectral attribute matcher is used to get a list of object types for a region. The experiment uses only regions obtained from the region segmenter. The system uses local schema regions to direct semantic merging, including adjacent regions. That is, it uses spectral properties and location. The result is to merge tree regions together, bush regions together, and grass regions together. The system uses long straight lines in 2-D schema; place a rectangular mask around selected schema edge. The system selects lines within the mask within tolerance of slope of schema line as candidates. It merges all colinear segments within the mask and matches all resulting segments to the schema line. It matches on slope, length, distance between center, and RMS error. In the example, the procedure matches three sides of the roof. The roof region now matches a parallelogram.

Symbolic Region Shape Matches via 2-D Schemas: Some schema regions have distinctive 2-D shapes, including trapezoid, rectangle, and ellipse. Properties for matching include size, aspect ratio, and color.

Overall results. The system does reasonably well in making a crude segmentation of the image. Tree, roof, grass, bush, shutter and sky are labeled appropriately. Major regions not labeled include much of the house, and trees in the background.

The perspective KS uses knowledge that bushes are vertical and stand on the ground plane, to estimate the range of the bush and its size. Bush adjacent to grass implies that bush is probably not occluded. Computed size partially validates identification as bush.

	spectral	texture	location	shape	size
tree	green	?irregular	above horizon above bush	tall	
bush	green	?irregular	below horizon above grass	low	
grass	green	?irregular	below horizon below bush	flat	
sky	light blue	?uniform	above all		large
house walls		?uniform	above and below horizon ?adj roof ?adj grass	?vert trapez	large
roof		?regular	above horizon ?adj walls	trapezoid str lines	large
shutter		?uniform	above and below horizon in walls	trapezoid	small known

Question marks indicate properties and relations that [Parma 80] do not use. They apparently do not use house walls adjacent to roof, house walls vertical, shutters symmetric, shutters inside house. They do not use occlusion or texture analysis.

The system has several interesting capabilities which belong in a general system. Its segmentation is limited. The quality of edges and regions holds back interpretation. Texture description is weak. Because it uses locations in an image, the system described is not only viewpoint-dependent but dependent on the specific scene. Their capabilities seem better than that. They imply that general vision can be achieved by having many special schemas and selecting among them. I disagree.

2.2.12 – Rubin 1978

ARGOS uses color, texture, adjacency, occlusion, location, size, and shape factors. It generates two dimensional models of the scene from various views. It attempts to generalize parts of the network.

Search is a form of dynamic programming with restricted transitions. Its first task is to find the view angle. Then it should name objects from known view. Without segmentation, its images were 75x100 pixels, with 7500 deep search. Experiments shown are with hand-drawn segments. They also use Shafer's version of Ohlander's segmentation. View angle is determined to 51 degrees.

ARGOS uses 'adjacency first-order' Markov evaluation which relates all surrounding nodes to the node under consideration. ARGOS has units called Primitive Picture Elements which may be segments from [Shafer 80], or which may be individual pixels. PPEs may be thought of as the largest region that is homogeneous in both signal and symbol. Both image models and test images are in the form of PPEs. ARGOS mostly relies on adjacency relations. For example, images of Pittsburgh will have mountains between sky and buildings. There may be vertical or horizontal adjacency nodes between PPEs. PPEs may have within relations, but without a containment hierarchy. All relations are single level and explicit. Networks tend to be large.

Much depends on spectral labelling. Median of blue gives 44% correct labelling. He ends up choosing median red, median blue, median green, contrast red, contrast green, and contrast blue, where contrast is from [Tamura 77]. Argos uses a weighted-Euclidean distance, weighted by an adjusted standard deviation. Statistics are obtained by computing mean and standard deviation for pixels in regions segmented by humans.

ARGOS uses locus search. It has a forward pass, keeping paths with values near the maximum value. Pruning heuristics are important. Since there is no unique order (two dimensional), paths are recombined, using maximum over neighbors. The pruning threshold is dynamic, relative to the best value at a given depth. Transition likelihoods are the network knowledge constraints. They have only three values, 0, .1, .9. Likelihoods at a PPE (pixel) are normalized, once computed.

In the backtrace, multiple beam pointers from the forward pass may disagree because the problem is two-dimensional. Some heuristics for conflict resolution were tried: throw a pixel away; carry along pointers when a pixel is left out; reject some possibilities based on adjacency rules; select by voting.

The internal model is a three dimensional model of the city which is used to generate all possible views. A network of relations (predominately adjacency relations) is constructed by multiple views. A region which appears in different views may be merged into one PPE, depending on adjacencies.

ARGOS does not segment, it labels. It works with pixels or regions. It can use absolute image location; i.e. mountains are usually found in the top of images. Each region has an MBR (minimum bounding rectangle) along horizontal and vertical axes. Proximity of MBRs is used to decide on merging two regions. Note that image location is used. Shape knowledge is difficult to incorporate, since segments may combine along self-transitions in the network. Four shape measures were used: fractional fill, compactness, orientation, and elongation.

ARGOS was found to work better without size knowledge (image size) than with size. No advantage was found in making a hierarchy of knowledge sources.

It discusses some extensions to other city scenes, to non-city scenes. None of these were implemented. Image knowledge should be divided into the scene level or schema level, viewpoint level, and object level. ARGOS does not address automatic model generation. It is assumed that all models are built by hand. He discusses a knowledge hierarchy to use more general schema, starting at the most general end of the hierarchy. He observes that lower levels of the knowledge hierarchy look alike.

Fifteen pictures taken from five different vantage points around Pittsburgh were segmented and labelled by untrained humans. Human labelling was to define ground truth. Seven pictures were chosen as a training set, the remainder as a test set.

The networks were reduced to 10%. Weights for terms were determined on the training set. The beam size was 25 entries. Data were smoothed in three ways: 1) simple smoothing, i.e. custering, changing labels of surrounded pixels; 2) throwing out any region with less than eight pixels; 3) filling unlabeled holes in regions.

The system was correct in 71% of labeled cells for the test set. They do not say what fraction are labelled. Hand-segmented data were used to determine view angle; rms error was 41 degrees. Accuracy of labels was 67%. Automatically segmented data gave view angle error of 60 degrees and accuracy of labelling of 59%. A priori, the system would not generalize for both close-up and distant pictures. Since it depends on horizontal and vertical dimensions, these are inversely proportional to distance. To include both, a very large range of dimensions would be necessary, which would make very weak dimension constraints. Adjacencies also change, e.g. buildings obscure mountains. Image location is not at all general. In summary, the system uses these types of information: spectral; adjacency; horizontal and vertical image dimensions; and absolute image location. Only spectral information is viewpoint independent. The system has little for a general vision system.

2.3 Applications

A few projections are made for applications of vision systems for the short term, for the mid-term (two to three years), and for the long-term (three to five years).

Industrial vision systems for the short term have a small market. Whether there is a profitable evolutionary niche or not remains to be seen. I believe that the major obstacle is not the lack of knowledge of users, although that is a factor, but the lack of capability of current industrial vision systems, which use a technology which is at least fifteen years old. They use thresholding to obtain binary images, two-dimensional models, and trivial global descriptors, like moments of the boundary. Now researchers are beginning to use local descriptors such as holes and corners to deal with obscuration. Adaptive thresholding is being developed to increase the flexibility of such systems, but that is greatly inadequate. As a consequence of thresholding, few features can be obtained, lighting must be carefully engineered, and the system is not rugged at all. In most cases, industrial vision systems won't work at all. Where they work, the applicability is limited and special case engineering is required, raising the cost to users, and the risk.

In the mid term, attaining larger markets is an important target, in going from an evolutionary niche to a major impact at the 10% level in automation. To achieve the target, systems must be much more capable, to lessen the extent of custom engineering. They must distinguish many kinds of flaws in inspection, especially cosmetic fr^m essential flaws. Systems can accomplish those objectives by incorporating several mechanisms including structured light, which begins to lead to three dimensional vision systems. Gray scale segmentation allows more internal detail, and less sensitivity to lighting. Greater speed and better shape discrimination will improve 2-D systems to a useful level.

In the long term, 3-D systems for warehousing, handling unoriented parts, and inspection of non-laminar parts will be important. The programming of vision problems will be a major issue, thus a single system will be essential for many applications. While a system may deal with only one object at a time, over a large company and over many companies, a system will require the generality to deal with many parts. Teaching and part programming will be difficult, requiring the ability to work from data bases of geometric models. Learning may also become important.

In cartography, current applications use automated stereo for terrain. These systems work only partially for terrain. Stereo mapping is very labor-intensive. As resolution requirements are increased by a factor x, the volume of effort increases by a factor of x-squared.

For the mid-term in cartography, automated stereo mapping for complex cultural sites is expected in research situations, and subsequently in production prototypes. Limited feature classification of linear features may be demonstrated. Aids for measurement of dimensions and data entry is expected to be important.

For the long term, automated feature classification is expected to make a major impact in mapping.

In photointerpretation in the mid-term, monitoring of selected objects in restricted situations, such as aircraft, vehicles on roads, and rail traffic is expected to be demonstrated to be feasible for classification, identification, counting and measurement.

In the long term, classification of a greater variety of objects with broader context and much greater detail is expected to be demonstrated.

Many other applications appear, some bordering on science fiction. They include guidance, medical image analysis for diagnosis, laboratory analysis, aids to handicapped. There are a number of hazardous environments in which robots with vision are expected to be used, including:

1. space craft servicing;
2. Communications is expected to be the dominant utilization of space. For antenna construction and servicing, the cost of maintaining a man in space is estimated to be millions of dollars, and in the radiation belt, there is no way of maintaining people for long times.
3. Undersea oil exploration, mineral exploitation, and naval operations;
4. Firefighting;
5. Maintenance in energy production has become prominent with the problem of servicing nuclear reactors, but the problems of power generation are generic. Similar problems are important for synfuel fabrication and fossil fuel power plants.
6. Electric line servicing has a high fatality rate.
7. Carrying tv cameras into hazardous environments for news;
8. The battlefield robot.

My own favorite is the home robot for cooking and cleaning. The home environment is complex. I don't expect to see home robots for decades.

There are important applications in psychology in determining perceptual mechanisms. Those mechanisms give a powerful insight into epistemology, in determining what people can know, the limits to human perception. They also have a potentiai for education, in determining how to teach people in ways that make greatest use of the natural perceptual mechanisms.

2.4 Objectives

The summary of applications leads to high objectives for a vision system: high performance; general; complete; intelligent; easy to use; and system support.

High performance relates to complex scenes with many objects, and applications requirements for detail, accuracy, resolution, and speed.

Generality implies generic with respect to object class, i.e. a system able to easy handle similar applications such as a mix of models of small electric motors. Generality also implies generic with respect to observation, i.e. viewpoint-insensitive and sensor-insensitive. A three-dimensional system is a primary means to achieve this objective. Generality also implies a standard system for research and applications, with a large, general core, plus mechanisms for special case implementations.

Completeness means spanning all applications tasks. This can be implemented by integrating all data, knowledge, and information, including multi-sensor data, expert knowledge, geometric models, prediction mechanisms, mathematics and physics. It also implies powerful perceptual mechanisms for observation, including strong use of shape, edge segmentation, texture and image organization, depth and stereo, shape from shading, and shape from shape.

Intelligent implies reasoning in the domains of images and surfaces, i.e. geometric reasoning, like a human observer or analyst. Mechanisms which make a system easy to use include standard user aids from computer science, an intelligent editor, automatic program synthesis, geometric models as a natural mode of communication common to man and machine, natural system structure to make the system intuitively clear to users at the system level, and bridges to natural language. System support assumes a large, critical mass effort, with continuing development, and a user base with collaborations.

2.5 System Design

Vision systems should integrate results of many image operators: region and edge segmentation; texture segmentation; surfaces from stereo, from motion parallax, and from object motion; shape from shape; surface interpolation; and shape from shading. Vision systems should combine such descriptions with general and domain-specific knowledge. Knowledge and observations relate geometric entities including: image; edges; structures of edges; surfaces; structures of surfaces; and objects. Thus, vision systems integrate inputs from many sources relating several geometric types. In computer systems terms, integration requires matching input/output formats of procedures, which becomes cleaner conceptually in terms of data types. But vision involves geometric types with well-defined transformation properties, not just data types. Representations for geometric structures provide the basis for integrating these different data and knowledge included in images, image structures, surfaces, and objects. The fundamental science base of vision systems is representation.

In building a vision system, whether through evolution or by design of a computer system, concern is given to eliminating duplication, choosing clean structures of data and control to eliminate inefficiencies of storage and computation, sharing data structures, and merging and streamlining data types. These efforts go to decrease the size of the system, biological or computer, increase its speed, and usually to simplify its conceptual and physical structure. The primary concern is achieving adequate perceptual performance within limits of computation structure and computation power. In the human, these limits are imposed by wiring limitations, size and power consumption. There are corresponding limits for computers. Adequate performance can be related to completeness of representation and completeness of perceptual maps, within complexity limits. These concerns for adequacy, simplicity, and efficiency are similar to those of the mathematician who is concerned with completeness and equivalence of mathematical types, or their simplest axiomatization, or who generalizes mathematical structures to treat many types uniformly, or the worker in analysis of algorithms who determines worst case computational complexity. Not only is adequate representation a means to making efficient and manageable systems, but adequate representation allows solving hard problems by compact restatement with constructs which expose inherent complexity rather than apparent complexity. Any program or biological system is a formal system. The concepts are defined; they may not be general, they may not be well-organized. The point is that analysis of generality and organization makes an important practical difference.

2.6 Representation

It is often useful to assume that biological evolution has had time for considerable optimization of the sort just described, well-structured representations. A good starting place for exploring biological perception is from the standpoint of representation.

Representation means different things to different people. This paragraph is a personal view. Much work deals with domain-independent representation, largely concerned with properties which are relevant to typed set theory, like inheritance. Set theory is inadequate for vision and probably inadequate for most domains of AI. Set theory or logic is general, but weak. Logic is valuable as a framework for embedding systems, but very little can be actually accomplished with such a weak theory. For vision, strong theories with limited domain are valid, e.g. topology and geometry. If geometric proofs and calculations are formulated in logic or in informal reasoning systems based on set theory, e.g. production rules, their statement becomes clumsy and proofs are very long, unless intermediate levels of mathematical theories are built up. Proofs are long and difficult because the language is weak. Mathematics is a compact and powerful language for expressing geometric concepts. It is important to represent these mathematical structures by building up an appropriate hierarchy of mathematical types.

Representation must concentrate on the problem domain and the task at hand. That does not mean that there is a separate representation for each domain and for each task. Physics does well with a small set of mathematical types and mathematical operations for solutions (a few hundred). Vector spaces, for example, are useful in many physics domains. Representation depends on the purpose, but a single representation may serve many purposes, and a few hundred representations serve all. There are few independent (inequivalent) mathematical structures and few independent (inequivalent) mathematical problems. Representations are open-ended and hierarchical, built from a few primitives and a few composition rules. Primitives at one level are compounded to become primitives at one level up. Applications tasks are built from a few fundamental mathematical tasks by composition. This promises a countable set of constructs, but not many have been constructed, and there are generalizing and compacting forces at work. Mathematics is concerned with equivalence of systems, an antidote to the process of inventing representations and names which turn out to be equivalent to other representations. Most representations are minor variations of others. Defining a neat hierarchy has largely been done by mathematicians.

In artificial intelligence, representation is a central concern. If representations are compounds, what are the fundamental representations which ultimately form their basis in a reductionist schema? I believe that the bases are fundamental mathematical, physical, and perceptual primitives. This does not mean mathematics and physics as taught in schools, but intuitive models which everyone apparently possesses to varying extent. Education does provide better models for some people, but on the whole, formal mathematics and physics depend on primitive, intuitive concepts. It is a personal belief that other domains are described by analogy with mathematics and physics. A current view is that language determines perception, a view largely discredited by experiment. The counter view is more tenable, that the representations of language are determined by perception.

Mathematicians use representation to mean a map from a co-domain A of some mathematical structure onto a domain of type B, a map which preserves the structure of A. Frequently the map is from an abstract to a more concrete type, for example, representation of the rotation group by the group of orthogonal matrices as transformations in Cartesian three space. This example is a homomorphism of a group onto the group of linear transformations of a vector space, hence matrices. More generally, a representation may map a group onto the transformation group of a vector space preserving composition. In artificial intelligence, the term representation refers to a map of any kind, without awareness of structure of co-domain and domain and without concern about whether the map preserves structure, even when there is obvious structure, and when standard mathematical terminology might be used.

Our central theme is **Structural Isomorphism**. One important representation is the shape of structural elements of objects relevant to a task. At a crude level, an artist structures a human figure in terms of torso, arms, legs, and head, and represents these parts as ellipsoids, or with slightly better approximations for parts. This level focusses on describing gross body shape in terms of part/whole structure and articulation. For a more lifelike rendering of individual parts, the artist focusses on muscles and bones, the structural elements of the parts. Muscles are laminae which flex and extend. Since there are only a dozen or so muscles per limb, the representation at this level of detail is only slightly more complex than at the coarser level, i.e. total complexity is only a small multiple of the complexity of the coarser level of detail. Representation of muscles is an obvious structure for the heart. At a tissue level, muscle fibers are natural structural elements. For biological objects, structure may be related to development. For manufactured objects, structural representations may be closely connected with fabrication and mechanical construction operations, which include milling, i.e. translation and turning, i.e. screw, extrusion, i.e. translation, and assembly operations, including insertion and screwing. Generalized cylinders are determined by the principle of generalized translational invariance, related to many such fabrication operations.

The importance of structural representations is: **Form equals Function**. Generalized cylinders were intended to provide suitable abstractions of shape to make compact, well-defined representations of function. It is a tenet that object classes are defined in two ways: by function and by abstractions from perception. Both definitions of object class are typified by abstract shape, i.e. generic shape elements and relations. Structural representations thus cover a very large part of relevant representation.

2.7 Criteria For Shape Representation

Criteria for 3-D shape representation were formulated in developing generalized cylinders and were described in [Thomas 74]. Generalized cylinders were initially intended for use in visual interpretation of complex objects as a means for a natural semantics for part/whole segmentation. The idea of a segmentation is not new, but the choice of primitive element determines whether the resulting segmentation into parts is useful. Also, they were intended for a compact representation of complex shapes from which symbolic relations between surfaces could be computed easily. The criteria for 3-D shape representation apply with appropriate changes for representing surface and image elements. The design criteria which led to the formulation of generalized cones were:

1) A representation of shape should aid in describing a very large possible class of objects, including many we have never seen. **The representation should be locally generated**, like splines from local primitives. It is not reasonable to enumerate rigid primitives like cube, sphere, cylinder, ... Part/whole segmentation describes one form of generation, but the primitives which go into the part/whole description must be locally generated. Volumes and surfaces should be determined from samples or boundary conditions, with interpolation and extrapolation constrained by general principles.

2) Defining senses of similarity is a central issue of perception. Each representation introduces a sense of similarity which is natural in the representation. Generalized cylinders were introduced in order to represent locally generated constructions from fundamental geometrical operations, e.g. sweeping [Binford 71]. Generalized cylinders were to be augmented by spheres which characterize constructions based on rotations. One interpretation of the phrase 'natural semantic' interpretation is in terms of these fundamental operations. A representation of shape should aid in describing similarities of classes of similar objects, i.e. it should be a generic representation. Each representation introduces a sense of similarity which is natural in the representation. Generalized cylinders were based on generalized translational invariance, which defines similarity in terms of a hierarchy of congruence transformations mapping one slice of a generalized cylinder into another [Binford 71]. Congruence transformations determine the set of similarities of parts, while part/whole relations define global similarity. Thus, in this framework, a stick is similar to a snake or a ring, and similar in another sense to a screw.

3) A representation of shape should aid in symbolic, generic prediction of appearances, generic with respect to object class and generic with respect to observation, i.e. valid over a broad range of viewpoints, illumination, etc. A representation of shape should aid in inferring volume description from image information, more generally, aid in symbolic, generic description, generic with respect to object class and with respect to observation.

4) A representation of shape should define levels of detail, coarse to fine, by defining a natural semantic segmentation, a part/whole decomposition intuitively natural to us. Parts should be part/whole structures themselves. This condition helps the system communicate with humans to help debug programs, and allows creating understandable systems. **Parts should be defined by continuity.** A surface is not a part in a 'natural semantics'; e.g. a cube has six surfaces, yet a cube is thought of by most people as a single part. If we define parts by surface continuity, then only separate objects are parts, and a man standing on a floor is not separate from the

floor. If, on the other hand, we define parts by surface tangent plane continuity, then a cube has six 'parts'. Parts are volumes. **Primitive parts should be generated from elements which are disjoint and for which a small, finite set gives a good approximation.** This adds intuitive clarity in description and in model building. Generalized cylinders correspond to stacking volume elements like slices of bread. Ribbon surfaces correspond to stacking surface elements. **A representation should be local.** If we want to describe an arm of a volume, we want to limit our attention to that part of the shape. Some splines have that property. The covering by a finite set implies that the elements are volumes. In the Blum transform [Blum 67] which is a covering by a minimal set of maximal interior disks, the elements are overlapping circles, i.e. not disjoint. In the Fourier representation, components or eigenfunctions form an overlapping set. Eigenfunctions for aircraft are roughly overlapping aircraft-shaped elements additively combined. It is much more natural to make disjoint combination of parts like fuselage, wings, and tail. The Fourier transform of a shape is global. Local changes have global consequences.

5) Parts should be locally realizable, i.e. they should be closed and non-intersecting. Surfaces must be investigated totally before closure and intersection can be tested.

We aim to represent elements by mathematical entities and to relate these entities by maps within levels and between levels. Several of the levels correspond to entities of three dimensions (volumes), two dimensions (surfaces), one dimension (curves), and zero dimensions (points) embedded in spaces of three dimensions, two dimensions and one dimension. We have maps which decrease dimension (projections) and maps which raise dimension (sweeping operations).

Our work has been based on the following paradigm. Descriptions are made of geometrical entities formed by two processes: a) grouping by a few geometric relational operations which are more or less independent of the geometric entity and which are common to all levels; these grouping operations correspond to neighborhoods of approximately uniform shape, elongated narrow neigborhoods in all directions corresponding to longitudinal projection, and transverse projection [Nevatia 77]; b) segmentation by tight constraints which are specific to the geometrical entity.

2.8 Interpretation

One paradigm for interpretation is template matching of images. This assumes image invariance and has little place in three dimensional image analysis. Template matching requires enormous computation for three space scenes; even that computational inefficiency is a min. fault compared to the fatal flaw: it produces a very weak sense of identity of objects. Another interpretation paradigm is graph endomorphism, finding an embedding of a model element in a description of an image, where both are expressed as a graph of nodes and relations.

A distinction is made between total and local representations and between total and local matching. A further distinction is made between arbitrary and semantic segmentations. Key issues are: 1) generic interpretation in terms of object classes, insensitive to viewpoint; 2) semantic interpretation, i.e. an evaluation function which accounts for details of the scene and observation process; 3) semantic search in matching, using semantic segmentations and indexing.

2.8.1 – Total Image Matching

Most recognition schemes rely on total image matching, complete image congruence. The problem posed is to match one image with another, invariant to translation and rotation; this assumes image invariance, i.e. there must be no systematic differences between images. Any systematic differences between images will result in large, unpredictable biases in location of best match. The limitations of total image matching schemes don't depend on whether they match intensities or Fourier Transforms or coefficients of other orthogonal expansions or eigenvalues of such expansions. These schemes sound more relevant than they are in reality, since they refer to plane figures as 'objects', which in plain English implies three-space elements. The usual approach is to store a dense set of possible views (or coefficients describing them) so that any sensed image is 'near enough' to one of the dense set of views. That rapidly becomes unmanageable. Consider three-space, articulated objects in arbitrary positions, orientations, and articulations. Assume that we don't know their shape exactly (the next person we see will be unlike any other). Objects are painted with irregular patterns which we don't know exactly either (people wear different clothes, different dogs have different spots). They will be in uneven illumination which we don't know exactly (depends on atmospheric conditions and reflections from nearby objects) with shadows and obscurations. Images depend on many things, even on position within the retina because of photometric and geometric non-uniformity (distortions and design). Sensors may have unusual response (SAR, infrared). Thus, store a dense set of views of all possible combinations of objects in all positions, orientations, and articulations, with all possible illumination for all possible sensors. Recall that any of these systematic effects can cause large, systematic errors in position of best match, and interfere with estimation of accuracy. The complexity for three dimensional objects with articulation and obscuration may be related to the power set of the plane with a certain granularity. We know a great deal about familiar objects and familiar scenes, but this is not the way to use that knowledge. Humans perceive scenes on postcards with ease without knowing surface pigmentation, illumination, or sensor characteristics. In a sense, humans are always seeing objects they have never seen before.

Even if total image matching could be made to work, it has little value, even for two dimensional scenes, because it gives a trivial sense of object. Distinct views of an object in three space are separate and unrelated in such a scheme, linked only by their object name. This is not a useful sense of object. In two dimensions, any plane figure is a separate figure, unrelated to any other. These schemes cannot identify shared image elements, i.e. partial image congruence, to perceive family relations.

2.8.2 – Local Image Matching

Some concepts are defined here:

1) Total representations of an image depend on the full image, e.g. Fourier Transforms of an image are defined on the whole image. Total representations over a domain are defined on the total domain.

2) Local or partial representations are defined on proper subsets of the domain represented, e.g. B splines are defined on a few adjacent nodes of a surface mesh. Locally generative representations like splines can be parameterized locally and composed globally to satisfy criteria such as continuity.

3) Semantically segmented representations of an image are local representations defined on image domains which are specific to image content. An example is a set of extended edges in an image. Semantically segmented representations of a domain (e.g. curve, surface) are local representations specific to the content of the domain. Another example is a approximation of curves by a B spline basis where nodes are chosen for optimal fit, corresponding to cusps. Local representations defined on an arbitrary tesselation of the domain are arbitrary segmented representations.

4) Part/whole representations are structured, semantically segmented representations with disjoint elements.

5) Matching is a map from two descriptions to a third description. The third description is usually a real number (distance or probability). In our work, we emphasize matches whose results are structured descriptions.

6) Local matching is a map from two local descriptions to a third local description.

One important generalization of interpretation is **local image matching of semantically segmented descriptions**, identifying shared semantic image components, especially those which are locally generative like curves and ribbons in images. Local representations may be used for total matches, e.g. least squares matching of all observed curves to all model curves. There is little gain. Local or partial matching allows identification in the presence of obscuration or physical differences between model and image. Some requirements of local matching include: a) sufficient support of local matching to generate adequate constraints; b) semantic segmentations to limit combinatorics of matching.

For physical reasons, curves are a useful structural element in images. Many image curves and groupings of curves correspond to surface edges and inherent surface markings. Some correspond to illumination discontinuities which are also interesting. Curves are also useful for describing many strictly two-dimensional forms, e.g. characters and drawings.

We suggest a principle for interpretation: **incorporate models with explicit representation isomorphic to the domain of variation.** In pattern recognition terms, this approach reduces cluster size by explicitly accounting for effects of 'nuisance' variables. If the actual domain of variation is a subset of a larger domain, then representations of the larger domain can be restricted, e.g. we can find simplified special cases such as viewing objects in a few stable states from restricted viewpoints. Partial image matching can relate common image structure of distinctly different figures, but interprets different poses of a three space object as different objects. From our principle we should represent three things explicitly in our models: three-space objects; the observation process; and illumination. Of course, if objects or viewpoint are restricted, then the representation is simplified. Use of image curves in image matching directly reduces some image variation caused by photometric effects, i.e. illumination, sensor differences, and sensor inhomogeneities. It is surprising how little we can do with image curves and structures of curves without three-space interpretation. Obscurations, surface markings, and edges of surfaces are three-space concepts. Image curve descriptions are compact and provide substantial reduction of combinatorics, however.

2.8.3 – Three-space Matching

In total three-space matching schemes, different articulations of a doll are separate objects, related only by their object name. Structured, partial matches do relate different articulations of an object as different aspects of the same object. While they seem gloriously general, partial three-space matches are not adequate, either. Our usual sense of object is not determined by a three-dimensional shape but by a class of three-dimensional shapes. While this may seem a hopelessly general paradigm for perception, we maintain that it is the usual perceptual problem and that mechanisms for generic interpretation are feasible. Interpretation by partial three-space matching would recognize different configurations of a 747 as distinct, unrelated objects, also different configurations of an F4 with different radar, fuel tanks, or armament, or a truck with different loads. It would have no idea that a 747B resembled a 747SP more than a truck. Without generic perception, there would be no cognitive basis for concepts of truck, vehicle, or passenger aircraft. Coffee cups vary from conical styrofoam throwaways to ceramic handcrafted treasures, from minimal forms to extravagant. Chairs have great variety. We have not seen all such possible forms, we cannot enumerate all possible prototypes, and if we see another which is distinctly different, we probably will interpret it correctly. Similarity, not spatial congruence, is the paradigm of interpretation in nature. As stated above, humans are always seeing objects they have never seen before. No one has ever seen me before as I am now (in the sense of spatial congruence); they may have seen me as I was on one occasion yesterday or two years ago. That is the central perceptual problem. The paradigm is recognizing a friend after ten years aging, ten pounds weight loss, in different clothes, with less hair, or recognizing a tiger from verbal descriptions and warnings about them, without having seen one before, even though no two tigers are identical. We have similar motivation for incorporating generic mechanisms for applications in manufacturing or photointerpretation to deal with the problem of programming a class of related tasks. Here, object classes may be generated by a variation of a design or manufacturing process, e.g. a small motor product mix. Variation may be small within one task, but among a class of tasks the objects form a class with considerable variability yet strong similarity.

One usual approach is to characterize a class as a prototype with a distance measure or as the union of such classes, or as some membership function on such sets, but that approach induces a weak sense of object class. Given fixed metrics based on three-dimensional distance for typical object classes we conjecture that we can choose elements of these object classes which are very far apart in the metric, such that if the class diameter is relaxed to include usual members of the class, then it fails to discriminate against non-members.

The essential definition of object class is functional. Manmade objects are designed for a function and living things have a teleology. Object classes have an associated three-dimensional form: form equals function. That is, an object's function is to be its form. Which aspect of its form? An object's function is often a geometric function. The function of a room is to be an enclosing volume. The function of a chair, desk or table is to be a flat surface at a comfortable height to sit, write, eat, etc. An object's function may constrain the choice of materials for its fabrication, e.g. a mattress vs. a desk. Cost and available fabrication processes may constrain otherwise free choices. For example, a runway has a minimum length constrained by its function in takeoff and landing of a class of aircraft, its maximum length constrained by cost. An important issue for interpretation is to identify which geometric parameters are causally determined such that their distributions are not just biases of the sample population, e.g. runway length. These essential characteristics enable tight discrimination which is not possible if properties are treated only as statistical distributions. Causal relations come from function. Such reasoning allows classifying elements into natural subgroups, e.g. runways classified by the types of aircraft they serve, even when insufficient statistical information or none is available. This capability is important since cases in which all distributions can be determined are probably few.

There is no great mystery about generic capability. The relevant geometric description may be at an abstract level. We must define object form from function in suitable abstractions, extract descriptions of form in abstractions, and find a set of equivalences. Because there are only a few descriptors of generalized cones and thus only a few levels of abstractions, this scheme is feasible.

Statistical approaches assume that the choice of features and class definitions may be given externally; they concentrate on the manipulation of a priori and observed probabilities. We are passionately concerned about just what defines specific classes and which specific features should be used. In the light of our paradigm, general vision requires strong description and weak classification, since a central proble n is defining object classes dynamically in a complex visual environment, i.e. creating new object classes while living.

Because of computational limits and information limits we have limits on perception. Computational limits imply that we compute only the simplest few of the enormously many possible functions on an image. Information limits imply that we can consider only relatively simple underlying object interpretations for observations. These models represent a commitment or preconception to perceive what we can within those limits. Observations are represented as instantiations of these models.

A purely statistical approach as opposed to a causal, structural approach has limits which follow from those information limits. The number of possible scenes far exceeds the number of measurements. The number of possible parameter combinations also far exceeds the number of possible measurements. Some simplifications are essential. Locality and decomposability into primitives are central in structural approaches. In a sense, generalized cylinders are singly curved, hence separable in internal coordinates. By contrast, the Fourier transform is separable in external coordinates.

2.9 An Approach To Interpretation

We offer one approach to the formalization of interpretation, to finding the 'best match' between an observation and models. Graph embedding is one approach: find an isomorphism of one or more models with a subset of the observation [Barrow 71]. This implies that models and observations have equivalent representations, e.g. comparing image models with image observations (the most usual), or comparing volume models with volume observations. The chief problem lies in defining criteria for inexact matching, for which the usual approach is syntactic, e.g. the best match has the fewest differences. This increases complexity of search greatly, but it also is a very unsatisfying sense of match. In some cases we expect differences, for example when limbs are obscured or not observable by the sensor used. Interpretation of significance of these differences is specific to the scene, to the observation process, to object class, and to articulation of the object.

Once a hypothesis has been made, predictions and measurements provide new information for the decision. Thus the decision is not necessarily made on a fixed data base, but interpretation becomes a problem of knowledge acquisition. This raises the concept of 'perceptual overkill'. It does not seem a useful heuristic to determine the minimal information for classification as an approximation to the maximum utility problem. In biological systems, many of the perceptual mechanisms are parallel and there is nothing to be gained in ignoring them. In machine perception, overwhelming verification of a correct hypothesis is typically inexpensive relative to computation required to get to the right hypothesis. These factors shift the utility balance toward getting data needed for a highly constrained decision. Very strong, relevant data are available if descriptive mechanisms can abstract them and interpretation mechanisms can use them. Object classes likewise have strong functional characterizations. A few structural relations characterize the class. Each relation can be tested.

The interpretation process is usually defined as a statistical process. Here, the definition is structural, detailed verification of criterial descriptions of class characteristics. All of this falls within the scope of decision theory, but usual applications of decision theory to interpretation trivialize the models from which conditional probabilities follow. The approach can be regarded as making explicit dependencies which might be glossed over in a statistical analysis. Noise is only one issue; systematic differences are primary.

The match of an observation to a model is the set of transformations necessary to map an abstract model to make it congruent to an abstracted. observed, semantically segmented description, both specified at the level of essential volume or surface elements. Multiple interpretations are resolved when possible by new observations as required. Essential and optional characteristics are confirmed by new observations so that all criterial structures are verified. The mapping of a structured match to a real number (probability) can often be postponed; at a final decision, a real number is usually convenient to characterize an acceptable subspace, but there are many other ways to characterize a subspace. The distance of the class of transformations can be evaluated in the semantic evaluation sense described above.

This approach is closely related to that of [Evans 68] in solving analogy problems. It also follows directly the fundamental definition of generalized cylinders by generalized translational invariance. Generalized cylinders are not defined by a distance function but rather by a congruence map. We pursued this approach of a metric on transformations in [Nevatia 74]. The outline of a simple distance measure can be specified based on some obvious semantics which are not complete or satisfactory. The distance measure is applied to the congruence transformation which includes articulation, scaling, rotation, obscuration, observation errors, object variations (growth, aging, missing parts, etc). The cost assigned to variations is highly context-dependent. Here are some examples. Articulations corresponding to usual postures and gaits have low cost. Articulations outside comfortable ranges have high cost. An obscuration interpretation must be consistent with observations of obscuring objects.

In summary, a key paradigm is similarity, not spatial congruence, matching objects which are similar but not identical. There are several important mechanisms:

1) description of three-space form in terms of generic shared structural elements and their abstractions;

2) inferring causally determined parameters;

3) characterizing object classes;

4) indexing subclasses of similar forms;

5) distance functions evaluated in context from congruence maps.

2.10 The Search Process

The search process in matching is important because potential complexity is high. Search for graph endomorphism has high complexity for graphs of moderate size. There must be semantic simplifications. We do not think it is reasonable in general vision to match against all models in visual memory. We introduced some concepts for indexing into subsets of objects similar to observed objects [Nevatia 74]. Those techniques were aimed at perception with a relatively large visual memory, even though we worked with only six objects. These indexing techniques relied on imposing a size (attachment) hierarchy on stick figures, i.e. on the part/whole graph. The sense of this was that small parts are attached to large parts. This led to comparing against similar structures, e.g. comparing a description of a doll with models of the class of objects which have two limbs at one end, three at the opposite end. Object classes were indexed by hash coding. We have considered a similar scheme in which object structures are arranged in a graph (based on topological and metric properties of stick figures) to be referenced by traversing the graph. These are mechanisms for generating hypotheses for subsequent verification. We focus on hypothesis generation because in vision it is a crucial step. In earlier work [Nevatia 74], an attachment hierarchy allowed structuring model graphs and description graphs, to facilitate indexing into a subset of similar shapes. Comparison of individual graphs was much abbreviated. This approach provides some capability for identification within a large visual memory.

It is a personal belief that general vision systems depend on building three dimensional descriptions, that prediction, description, and interpretation take place largely in three dimensions. A recent article describes constraints which lead in the direction of incorporating these capabilities in ACRONYM [Binford 81].

One paradigm for intelligent systems is prediction-hypothesis-verification. The paradigm can be useful or ineffectual depending on how adequately prediction, generating hypotheses, and verification are conceived. We refer to hypotheses as descriptive maps or cueing or bottom-up maps, i.e. mapping upward among structures in a geometric hierarchy, from image curves to surface edges, from structures of curves to surfaces, from structures of surfaces to objects. In the same way we refer to prediction maps or top-down maps. We normally think of vision hypotheses as great leaps from images to objects, and predictions as great leaps from object models to images. Those are not useful starting points. Vision systems have primarily been built on this basis, with only two levels of representation, name level and image level. Image level includes observations extracted from images and appearances of objects. This is too shallow a geometric structure. Instead, we have long thought of hypothesis-prediction-verification loops as steps between any two levels in a well-defined geometric hierarchy, a hierarchy which is deep with small steps between levels. Hypothesis-prediction-verification especially relate nearby levels, from image curve elements to extended curves to structures of curves (image organization, one level of Gestalt organization), to surfaces, through levels of surface organization, etc. Each loop of prediction- hypothesis-verification is at once bottom-up and top-down. We claim an essential unity of bottom-up and top-down operations. If strong context and weak visual data are available in a situation, a system appears top-down; if the system has strong visual data and weak contextual information it appears bottom-up; usually both context and visual data are available. Prediction is as effective at low levels of the geometric hierarchy as at the high level. Prediction is as general as description. If we have weak contextual information, if we don't know what objects are in the scene and we don't know our viewing conditions, we still know that we can represent objects by locally generative shape primitives, e.g. part/whole graphs of generalized cylinders, and we know that we can make generic predictions of the appearances of generalized cylinders, predictions which are insensitive to viewpoint, illumination, and sensor, general conditions on image areas and image curves corresponding to surfaces, limbs, and edges of surfaces. We remarked above that prediction-hypothesis-verification loops link primarily nearby levels in the geometric hierarchy. The nearer the levels, the more plausible the hypothesis. In ACRONYM, cueing (i.e. hypothesis generation) is based on powerful shape descriptions, image ribbons, surface ribbons, and generalized cylinders. We originally devised generalized cylinders as a natural way to use image cues about surfaces.

2.11 References

[Ballard 78] Ballard, D., C. Brown, J. Feldman; "An approach to knowledge-directed image analysis," in **Computer Vision Systems**, A. Hanson, E. Riseman, eds, Academic Press, New York, 1978. (cited on p. 65)

[Barrow 71] Barrow, H.G., A.P. Ambler, R.M. Burstall, "Some techniques for recognizing Structures in Pictures," *Int. Conf on Frontiers in Pattern recognition*, Honolulu, Hawaii, Jan 1971. (cited on p. 84)

[Binford 71] Binford, T.O., "Visual Perception by Computer," invited paper at the *IEEE Conference on Systems, Science and Cybernetics*, Miami, December 1971. (cited on p. 79)

[Binford 81] Binford, Thomas O., "Inferring Surfaces from Images," *Artificial Intelligence*, vol. 17(1981) 205–244, August 1981. (cited on p. 37,62,86)

[Blum 67] Blum, Harry, "A transformation for extracting new descriptors of shape," *Symposium on Models for Perception of Speech and Visual Form*, 362–380, ed. Weiant Whaten-Dunn, MIT Press, Cambridge, Mass., 1967. (cited on p. 80)

[Bolles 76] Bolles, R.C., "Verification Vision Within a Programmable Assembly System," AI Lab, Stanford University, Memo AIM-295, 1976. (cited on p. 66)

[Brooks 81] Brooks, Rodney A., "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial Intelligence Journal*, vol. 16, 1981. (cited on p. 34,35)

[Evans 68] Evans, Thomas G., "A heuristic Program to Solve Geometric Analogy Problems," **Semantic Information Processing**, ed. Marvin Minsky, MIT Press, Cambridge, Mass., 1968. (cited on p. 85)

[Faugeras 80] Faugeras, O., and K. Price, "Semantic Description of Aerial Images Using Stochastic Labelling," *Proc ARPA Image Understanding Workshop*, 89, Univ of Md, April 1980. (cited on p. 67)

[Garvey 76] Garvey, T.D., "Perceptual Strategies for Purposive Vision," SRI AI Center Tech Note 117, 1976. (cited on p. 67)

[Hewitt 68] Hewitt, C., "Planner," MIT AI Memo 168, 1968. (cited on p. 50)

[Horn 73] Horn, B.K.P., "On Lightness," MIT-AI Memo 295, 1973. (cited on p. 36)

[Hueckel 73] Hueckel, M., "A Local Visual Operator which Recognizes Edges and Lines," *J.Assoc Computing Machinery*, 20, 634 (1973). (cited on p. 67)

[Land 77] Land, E.H.; "The Retinex Theory of Color Vision," *Scientific American*, 1977. (cited on p. 36)

[Levine 78] Levine, M., "A knowledge-based computer vision system," in **Computer Vision Systems**, A. Hanson, E. Riseman, eds, Academic Press, New York, 1978. (cited on p. 68)

[Lowe 81] Lowe, David G., and Thomas O. Binford, "The Interpretation of Geometric Structure from Image Boundaries," *Proc. ARPA Image Understanding Workshop*, 39–46, April 1981. (cited on p. 62)

[Miyamoto 75] Miyamoto, E., and T.O. Binford, "Display Generated by a Generalized Cone Display," *Proc. Conf. on Computer Graphics, Pattern recognition, and Data Structures*, May, 1975. (cited on p. 67)

[Moravec 77] Moravec, H.P., "Towards Automatic Visual Obstacle Avoidance," *Proc 5th IJCAI*, 1977. (cited on p. 66)

[Nagao 78] Nagao, M., T. Matsuyama, Y. Ikeda; "Region Extraction and Shape Analysis of Aerial Photographs," *Proc 4ICPR*, p 620, 1978. (cited on p. 34,35,37,40)

[Nagao 80] Nagao, M., and T. Matsuyama, **A structural analysis of complex aerial photographs**, Plenum, 1980. (cited on p. 37,41)

[Nevatia 74] Nevatia, Ramakant, "Structured Descriptions of Complex Curved Objects for Recognition and Visual Memory," Ph.D. Dissertation, Dept. of Computer Science, Stanford University, STAN-CS-74-464, October 1974. (cited on p. 85,86)

[Nevatia 77] Nevatia, R., and T.O. Binford, "Description and recognition of Curved Objects," *Artificial Intelligence Journal*, 1977. (cited on p. 80)

[Ohta 80] Ohta, Y., "A region-oriented image-analysis system by computer," Thesis, Dept of Information Science, Kyoto University, 1980. (cited on p. 34,35,35,36,48,49)

[Parma 80] Parma, C.C., A.M. Hanson, E.M. Riseman; "Experiments in Schema-Driven Interpretation of a Natural Scene," Univ of Mass COINS Tech Rept 80-10, 1980. (cited on p. 70,72)

[Rubin 78] Rubin, S., "The ARGOS Image Understanding System," *Proc ARPA IU Workshop*, Nov 1978; also "The ARGOS Image Understanding System," Ph.D. Thesis, Carnegie-Mellon University, 1978. (cited on p. 35)

[Shafer 80] Shafer, Steven A., "MOOSE. Users' Manual, Implementation Guide, Evaluation," Bericht 70, Report IfI-HH-B-70/80, Fachbereich Informatik, Universität Hamburg, April 1980. (cited on p. 73)

[Shirai 78] Shirai, Y., "Recognition of man-made objects using edge cues," **Computer Vision Systems**, A. Hanson, E. Riseman, eds, Academic Press, New York, 1978. (cited on p. 34,35,63)

[Tamura 77] Tamura, H., S. Mori, and T. Yamawaki, "Psychological and Computational Measurements of Basic Textural Features and Their Comparison," ETL and Waseda University, Japan, 1977. (cited on p. 73)

[Thomas 74] Thomas, A.J., and T.O. Binford, "Information processing Analysis of Visual Perception: A Review," Stanford Artificial Intelligence Laboratory, AIM-227, June 1974. (cited on p. 79)

IMAGE SEGMENTATION

3.1 EDGE DETECTION AND REGION GROWING

3.1.1 – Why edge detection?

Edge detection, in the form of spatial differentiation, appears in the computer vision literature as early as 1955 [Dineen 1955]. This early and sustained interest arises from a perception that the types of patterns significant to a visual system consist of approximately homogeneous regions separated by abrupt boundaries. Although years of experience have shown that real digitized scenes are not easily characterized this way, the idea has persisted tenaciously, for the following reasons.

First, from an *introspective* point of view, one tends to believe the world to be composed of *objects*, each homogeneous in its cohesion, and abruptly separated from other objects and the background. That is an essential aspect of our way of perceiving the world, pervading disciplines from anatomy (where every bump, nodule, fascia, and tissue type is seen as a separate structure) to quantum mechanics (in which an essentially continuous all-pervasive field is seen to describe a separate localized particle). Whether this discretization of the world is a part of the structure of the world or of ourselves we cannot say (arguably it is impossible to say); nevertheless, it is here to stay.

Now close inspection of digital images, or for that matter, paintings from past centuries, leaves little doubt that the image of a single (realistic looking) object can but rarely be described as "homogeneous." Yet, even upon making such an observation, one's natural way to describe such a heterogeneous object is still to partition it into homogeneous components. The tendency is always to subdivide, perhaps reflecting the reductionist ethos widespread in modern science. One might expect that in artificial intelligence jargon, this process would be called the *chotomizing heuristic*; in fact, it is called *segmentation.* (Or borrowing from Caesar, subdivide and conquer.)

The products of the subdivision are homogeneous entities. For a human, the homogeneity is one of description, while for the machine it is generally one of measurement. Now, a measurement is a description, and a set of descriptions is a description, so we have to explain what we mean by these terms a little more precisely. By a description, we mean something which might be quite complicated from a machine perspective, encompassing such explicit descriptions as "it gets darker and redder from right to left, with a speckling that looks like that on a trout, but which fades into a very dense network of lines in the periphery." That would not generally be found to be a homogeneous region by a machine. A measurement, on the other hand, is meant to connote something very close to the language of the transducer providing the image data, e.g. brightness or range values. Most of the definitions of homogeneity implicit in automatic segmentation programs stray little from a constancy of such a measurement, though the situation appears to be improving.

By demanding a ***homogeneous*** description to define a homogeneous region, we mean that the description can have no ***explicit*** mention of boundaries or constituent regions. The relatively weak condition of explicitness is right because an ***implicit*** boundary would not be a property of the description, but rather something inferred from it. Thus a description such as "the value goes linearly from 100 in the lower left to −100 in the upper right" would be judged homogeneous, despite the well defined diagonal boundary separating positive from negative values. For the time being, we are content to include as homogeneous such descriptions as "the intensity goes as a step function..."

In more familiar terms, the description is nothing more than a model of the data in some system of ***representation*** of the pertinent knowledge. Including boundaries or regions per se in the representation language makes vacuous the notion of homogeneity. Our observations above are merely rewordings of the thesis that humans have a very rich representation language available, while machines as yet do not. Incidentally, the term ***representation language*** is meant to refer to the internal representation, whether it be essentially symbolic, continuous, or whatever, and shouldn't be confused with the language we use to communicate about the aspects of the representation introspectively available to us.

The problem of finding homogeneous regions can be approached either by finding the regions directly, whose difficulty increases with the complexity of description, or by finding the boundaries between the regions. Thus the first motivation for edge finding is for the purpose of segmentation into homogeneous regions in accordance with our own models of the world. It is based on ***introspective*** observations.

A second motivation is derived from ***extrospective*** observations which have been made by physiologists, perception psychologists (not to be confused with perceptive psychologists), and perceptual psychophysicists. Anatomical structures have been found which respond to abrupt changes in intensity and color as functions both of position and time. Observations have been made that indicate absolute colors and intensities, as well as their slow changes are not readily perceived, but abrupt changes are. Whether it is wise to mimic nature, or rather to attempt to mimic the precious little we think we know about nature, is problematic, despite the widespread tacit acceptance of the idea. It is worth considering that we are unlikely to find physiological processes involved with things that are not already a part of our introspective models. For example, we look for evidence that the visual system performs Fourier transforms, since Fourier transforms have a particular intuitive appeal. But they can also be viewed as only one of a myriad of possible isometries of a function space, special because they transform convolutions to multiplications. But that special property is mainly significant for ***linear*** systems. The theory for nonlinear systems is not as well developed, and not so widely known. Who can say that there is not some other isometry that is more appropriate for a nonlinear system? But how is a physiologist or psychologist to seek evidence for such an unknown object?

Finally, a third motivation for edge finding is based on the ***computational*** consideration of efficiency. Since boundaries are of a smaller dimension than images or regions, they are easier to handle: e.g. (for a smooth boundary, rather than a fractal) if the number of boundary points increases as $O(n)$, with $1/n$ the discretization interval (grid size), then the number of region points increases as $O(n^2)$. Also, the 1-dimensionality of a boundary provides a natural ordering for its points, which is easily translated to a processing order for a sequential algorithm. While the entirety of an image is filled with region points, only a small fraction constitute edge points.

The sparsity of edge points among image points is a major attraction of edge detection as an early step in stereoscopic vision, since it extremely diminishes the size of the search involved in matching points between the two images. Of course, this is only useful if the edge points bear some relation to fixtures in the world, so as to vindicate the assumption that edge points must match edge points. As it happens, this seems in fact to be a good assumption, and edge points appear to be more stable than more rudimentary features of intensity, such as the actual brightness values.

So far we have given a very general definition of edge detection as finding the geographical limits of a description. It is probably fair to say that few if any authors of edge detection methods thought that was what they were doing. The universal goal of edge detection algorithms is to find places in the image which a human would classify as "edges" or "boundaries." We apologize if that seems a trivial statement, but since we do not know how a person segments a scene, we are in no position to give an authoritative definition of what constitutes an "edge." Everyone agrees that a transversely translated step function ought to be called an edge. This corresponds to a boundary between regions of uniform intensity measurement, uniform at least near the boundary. Very little attention has been paid to any other definition of "edge," despite the fact that close observation of images reveals that step edges between uniform (strip) regions are exceedingly rare. This is not to say that edge detectors built to detect step edges don't find real ones; indeed they do. But it would be easy to believe that in most cases it is more by good fortune than design.

The term "edge" has been fairly widely abused, and we will continue that tradition here. One type of edge is that resulting from the boundary of some object. There are also edges which are merely boundaries between surface features. There are *local* edges and *global* edges, which are frequently called *contours*. Local and global are relative terms, and we mean them in comparison either to image or grid size. A local part of a curve, for example, would be well approximated by a straight segment in the given grid size. Thus another way of looking at the difference between local and global is related to manageable and unmanageable search problems, since locally all possible curves can be represented as all possible line segments on a coarse grid, while globally the space of all possible curves is vast.

3.1.2 – Local edge detection

We will not attempt to give a mathematically precise as well as operationally general definition of "edge" here. Properly, to do so one would study the imaging process as well as real images. Herskovits and Binford [Herskovits 1970] did so to a limited extent, presenting essentially 1-dimensional results. Essentially, what people have been looking for as edges are places with a large gradient, or places which resemble a step function in cross-section. It turns out that a number of different outlooks on how to look for these features lead to essentially the same computational technique, viz. convolution with some kernel followed by thresholding. (Strictly speaking, it is usually cross-correlation which is implemented, but since the families of kernels involved are complete under inversions, we take liberties with the term "convolution.")

Spatial differentiation and gradient estimation

If edges are places where things change fast, then the obvious way to look for them is by performing a spatial differentiation. This may be done by some discrete analog of the gradient, which is implemented by convolving with a kernel of small support. The smallest possible support for a differentiation is 2 pixels, and in such a case the convolution is often thought of as taking adjacent pixel differences, or first differences. Larger supports allow more creativity in the choice of the convolution kernel defining the differentiation, and provide the benefit of improved noise behavior. A great many authors estimate gradient or "stepness" by computing adjacent pixel differences. Martelli [Martelli 1972, Martelli 1973] and Turner [Turner 1974] are examples of the latter. Another way to think of the gradient is as a derived parameter of fitting a plane to the data. For sufficiently symmetric supports, this can also be implemented as convolutions. In fact many outwardly sophisticated techniques have.as their core the estimation of gradient.

Template matching and matched filtering

A popular way to look for features is with a matched filter or template, and this is quite common for step edges. Again the cross-correlation with the template, or the space domain realization of the filter are implemented as convolutions. The idea is that the "template" (the convolution kernel) is an ideal case of the feature one is seeking, and one looks for large values of the correlation as indicating the presence of the feature. Implicit in the use of the term "template-matching" is the implication that no attention is being paid to the vector space projection analysis of the process. Examples are the operators of Sobel [Duda 1973] and Kirsch [Kirsch 1971], as well as many others (further examples can be found in [Abdou 1978] and [Rosenfeld 1976]). The matched filter approach is operationally the same, but includes the analytical idea that as a consequence of the Cauchy-Schwarz inequality, the maximum response for normalized data occurs when the data is the (complex conjugate of the) template. Duda and Hart [Duda 1973] provide a more detailed discussion of the ideas of spatial differentiation, gradient estimation, and template matching, with a slightly different viewpoint. Shanmugam *et al.* [Shanmugam 1979] seek a slight generalization of the matched filter, in the sense that the filter must be strictly bandlimited and the objective is to maximize the power of the step response in a given space interval.

Locally, i.e. at a single point of the convolution result, the integration against the kernel can be thought of as orthogonal projection onto a 1-dimensional subspace of $\mathbf{R}^n$, where n is the number of pixels in the support of the kernel, and the projection is with respect to the usual inner product on $\mathbf{R}^n$. If there is more than 1 subspace involved, i.e. more than one convolution, then one has components which can be thought of as components of a vector in the space spanned by the subspaces. Then one can compute a magnitude for that vector (so as to get a number representing "edgeness" for thresholding). The magnitude may be in the Euclidean norm

$$\|x\| = (\sum v_i^2)^{1/2},$$

or in some other norm, such as the max norm

$$\|x\| = max\{v_i\},$$

or the sum norm

$$\|x\| = \sum |v_i|.$$

Best edge fit and optimal estimation

The simplest edge model, a translated step function, has 3 parameters (for a 2 dimensional picture). These might be, e.g., angle, left height, and right height. With enough normalization, these can be reduced to the single parameter of angle. Template matching methods use a separate template for each angle considered. But one can also try to determine the angle that best accounts for the data. Furthermore, the model may have more parameters, and there may be statistical information available.

The simplest type of best fit problem occurs when the model space is a linear subspace of the data space, which is an inner product space. In that case, the best fit is obtained by orthogonal projection to the model space. This is a very common method for fitting functions in 1 dimension, based on the observation that translation in space is equivalent to multiplication by a complex valued function of frequency in the frequency domain, so that all the translates of a given frequency component make up a linear subspace. In 2 dimensions, though, matters are complicated by the presence of rotations, so that while the same artifice applies to translations, the Fourier equivalent of rotation is still rotation, and the set of all rotations of a component is no longer a (1-dimensional) linear subspace, so direct orthogonal projection is no longer applicable. Hence many methods which seem very clever for 1 dimension fail for 2 dimensions. However, this nice property of Fourier transforms for 1 dimension can be thought of as a special case of a more general principle, which may be of use in inventing best fit methods. Specifically, one way to restate the spectral theorem [Halmos 1957, Halmos 1963] is that any normal operator in a Hilbert space is unitarily equivalent to a multiplication. For our purposes the Hilbert space can be taken to be $L^2(\mathbf{R}^2)$. Then the spectral theorem can be interpreted to say that given a normal operator A, we can find some isometry $U : L^2 \to L^2$ and some function $\varphi \in L^2$ such that $U^{-1}AU(f) = \varphi f$ for all f simultaneously. If A is a translation operator, the Fourier transform is such a unitary transformation, as we mentioned above. According to the theorem, there is some isometry of $L^2(\mathbf{R}^2)$ which will transform rotation into complex multiplication. Using that isometry like a Fourier transform, one could use projection methods to find best fits. Even better would be a transform that worked for translations and rotations at the same time, but that is impossible because translations do not in general commute with rotations (as would clearly be necessary for the existence of such a transform because multiplication is commutative).

A slightly more general best fit problem occurs when the model set consists of an n-parameter family of functions, and the object is to find a member of the family which minimizes some error measure with respect to the datum. If the family is differentiable, it can be thought of as a submanifold of the ambient space. Frequently the error measure is a metric on the space, and then the problem is seen as one of finding the closest point of the model manifold to the datum. In the case of estimation, there is a probability distribution involved, and one seeks a set of parameters minimizing the *expected* error.

Altes [Altes], Hueckel [Hueckel 1971,Hueckel 1969], O'Gorman [O'Gorman 1976], Abdou [Abdou 1978] find best fit edges. Altes uses essentially the 1-dimensional Fourier method described above. Hueckel and O'Gorman minimize the distance between the projection of data and parametrized model onto a truncated orthonormal basis, deriving the "optimal" parameters. However, both the number of parameters and the number of terms in the series are too small to allow good performance. Altes uses a more realistic edge model (in 1 dimension), but his results are not readily generalized to 2 dimensions. Abdou finds the best fit edge by what is essentially an exhaustive search over a slightly more general but still too simple model space, namely linear ramps between constants.

When the parameters one is seeking are the coefficients in an orthonormal basis, the parameters can be obtained simply by taking the inner product with the basis elements.

Higher order derivatives

Methods that rely on estimates of the gradient, or whose response is largely determined by the gradient cannot distinguish smooth transitions from abrupt ones. In the Hueckel and O'Gorman approaches, for example the early term(s) in the expansions are essentially the gradient. One approach to this problem is to use a preprocessing step which takes linear functions to 0. This idea is advanced by Binford [Binford 1981] in the form of "lateral inhibition," and in fact operators modelled on second and higher order derivatives will have this property. (It's interesting to consider just how many such operators there are. Suppose the operator support is n pixels. There are then n linearly independent such operators. Requiring that all operators take constants to 0 is a linear constraint, and that they take all linear functions to 0 is 2 more linear constraints, so there are $n-3$ linearly independent operators fulfilling the constraints. One may impose further onstraints by requiring various symmetries, and each discrete symmetry will reduce the dimension of the operator space by 1. For large supports, it is clear that there are many candidate operators.) The second derivative in the calculus of several variables is the Hessian, which is a matrix. Its algebraic invariants are the geometric invariants of the original function viewed as a surface. Various combinations of its components (taken linearly and nonlinearly) can be used as 2nd derivative operators. If an edge is sought at suitably defined maxima of the gradient, then for a 2nd derivative operator, one seeks 0-crossings. Marr and Hildreth [Marr 1979] use an approximation to the Laplacian, which is the trace of the Hessian. Dreschler and Nagel [Dreschler 1981a, Dreschler 1981b] use the determinant of the Hessian. Beaudet [Beaudet 1978] computes rotationally invariant derivatives up to 4th order.

Approximation and representation of image function

One of the drawbacks of the methods we have been describing is that a very few parameters are derived by some kind of local projection. The parameters are chosen for semantic interest, but while they respond well to intended features, the same is often true for unintended features. We have the following situation. Let X be the space of all local images, and $F \subset X$ the features one is seeking. Perhaps this is done by some map $\varphi : X \rightarrow \mathbf{R}$. One designs this map so that $\varphi(F) \geq \Theta$, for some threshold Θ, and one would like to be able to infer that if $\varphi(x) \geq \Theta$, then $x \in F$. Clearly, to do this, one must have some information about $\varphi^{-1}(-\infty, \Theta)$, but this is surprisingly often neglected.

Another way to think of this is that the few semantically derived parameters actually do not provide enough information to understand the structure of the image intensity function, even locally. Now, of course the pixel values constitute complete information, but it is not directly usable. One approach, then, is to seek a local representation for the image data which is appropriate for the questions one wishes to resolve with further processing. Approximating the Hessian is such a process, since that can be regarded as finding the best local quadratic approximation, just as computing a gradient can be viewed as finding a planar approximation. Prewitt [Prewitt 1970] computed her gradient parameters based on a planar fit. In the same vein, Haralick [Haralick 1980] fits planes to the data and defines edges as boundaries between maximal domains of fit, relative to an error measure. Planar fitting is very crude so he [Haralick 1981] proposes polynomial fitting as an extension. Beaudet [Beaudet 1978] is motivated by fitting a truncated Taylor series, though the semantics he ascribes to his operators are somewhat naïve. Hsu *et al.* [Hsu 1978] use a quadratic fit, based on Beaudet's techniques. Altes' work [Altes] is put forward as essentially a spline fit.

3.1.3 – Global edge detection

The Hough transform [Hough 1962, Duda 1971, Duda 1972, Duda 1973] is a technique to find collinear sets of feature points over an entire image. This can be applied in complete globality, i.e. over the entire image at once. Ballard and Sklansky [Ballard 1976] Shapiro[Shapiro 1974, Shapiro 1975, Shapiro 1978], and others use generalizations of the method to look for other 1 dimensional objects.

Frequently, the term *linking* is used synonymously with global edge detection. Linking consists of making lists of local edge elements connected head to toe, each list corresponding to an extended (global) edge. This is the most common global edge detection method, dating back at least to Roberts' work [Roberts 1963], and including many others [e.g. Horn 1972, Binford 1970, Nevatia 1978]. These methods differ primarily in the predicates used to determine whether to join a particular edgel into a contour. A major difficulty stems from the fact that the linking proceeds only after irreversible decisions are made about local edges, e.g. limiting each pixel to having an edgel of unique orientation, or making a binary decision about the presence of a local edge. The type of information available to the linking, which generally proceeds locally, is inadequate for many situations.

An improvement on the linking method is advanced by Montanari[Montanari 1970, Montanari 1971] and Martelli[Martelli 1972, Martelli 1973]. Here the prior local commitment is less extreme, and dynamic programming or heuristic graph search methods are used to find optimal paths with respect to some figure of merit. The figure of merit, a global parameter, replaces the local predicate as the contour selection method, and likewise as the main artistic element.

The "relaxation" methods propounded by Zucker, Rosenfeld, et al. [Zucker 1977, Rosenfeld 1975] attempt to find the contours globally, in parallel, and without excessive initial commitment. The process depends on a local pairwise reinforcement-inhibition process between edgels. The art is in choosing the reinforcement process. Explicit global edges are not produced, but presumably the process terminates with sets of edge points which are both connected and of a desired minimum length, which are then readily identified.

3.1.4 - Region growing

We motivated edge detection as a means to region finding. Why not just find the regions directly? Many people have tried doing just that. The advantage is that one is dealing with a global object, so the problem of linking is (or seems to be) avoided. Rather than deciding whether an edge separates 2 points, one must decide whether 2 points belong to the same region. Seen thus, the difference is mainly one of (linguistic) semantics. The data structures reflect regions, not edges, as do the algorithms. Consequently, despite the conceptual equivalence with edge finding, different approaches, harder to express in the edge detection paradigm, are developed. The simplest method is based on segmentation simply by intensity or color value. Brice and Fennema [Brice 1970, Fennema 1970] take this as their starting point, and then try heuristics to clean up. Ohlander [Ohlander 1975] segments based on dividing bimodal histograms of several color parameters. Shafer [Shafer 1980] builds on Ohlander's work. Somerville and Mundy [Somerville 1976] use a technique based on more sophisticated reasoning. They grow regions based on the uniformity of an approximation to the normal to the image intensity function. Kirsch [Kirsch 1971] defines regions based on thresholding a "contrast" (gradient) function.

3.1.5 – Statistical methods

Simple statistical models have been used in both edge detection and region growing. [Yakimovsky 1976] uses standard hypothesis-testing techniques to determine whether adjacent sets of pixels are samples from the same Gaussian distribution. [Griffith 1973] develops a detailed model of images of prismatic solids and computes the conditional probability that, given image data in a narrow band, the data contain a vertical line or edge. Chen and Pavlidis [Chen 1980] segment the entire image into regions uniform in the sense that each has been found to be a collection of random variables from the same Gaussian distribution. [Cooper 1980] take a maximum likelihood approach to find the boundary of a blob of a constant grey level on a background of constant but different grey level, with gaussian noise superimposed on both; both the "squiggliness' of the boundary and the intensity distribution within the two regions it implies affect the likelihood of the image data. Macrenhas and Prado [Macrenhas 1978] combine a somewhat more sophisticated image model with hypothesis testing to test square neighborhoods of four pixels for the likelihood of all possible edges.

Machuca and Gilbert [Machuca 1981] propose a "moment operator" for detecting edges and compare its performance to the Sobel and gradient operators in a variety of theoretical and experimental tests.

3.1.6 – Curve Segmentation

Rutkowski and Rosenfeld [Rutkowski 1978] propose several *ad hoc* methods of detecting angles or "corners" in digital curves and compare their performance with each other and with human performance on a test curve. [Shirai 1975] segments linked edge data into lines and curves based on a discrete curvature measure and then fits lines or curves to the data as an intermediate step in real-world object recognition. [Duda 1973] describe a method of curve segmentation called iterative endpoint fitting. [Pavlidis 1977] contains an extensive discussion of fitting continuous functions to discrete data: approximation, interpolation, splines, and some nonlinear methods are covered, and an extensive bibliography is included.

3.1.7 – Summaries of individual publications

Following are commentaries on a number of works in segmentation. The list is by no means exhaustive, but is intended to include the most influential works as well as some other representative research.

Abdou 1978

Abdou,Ikram Escandar; "Quantitative Methods of Edge Detection," Ph.D. thesis, University of Southern California, July 1978. Also USCIPI Report 830.

This work is concerned solely with local operators.

The author presents a review of several such operators:
Roberts
Sobel
Prewitt
Compass gradient
Kirsch
3 level, 5 level
Hueckel

It is interesting to note (as a comment on the literature in general), that he presents 8 different 3x3 convolution operators. There can be only 9 linearly independent such 3x3 operators (since they make a 9 dimensional vector space). The 8 presented are in fact linearly independent, and the addition of a single pixel operator, e.g.

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

would span the entire space.

He evaluates the performance of convolution operators on perfect step edge visual input, assuming square pixels and area proportional sampling (i.e. the pixel value $g(p)$ is defined by

$$g(p) = \int_{R(p)} f dA,$$

where $R(p)$ is the (square) pixel support region). This leads to complicated formulae for pixel values from rotated edges.

He discusses statistical aspects of edge detection and evaluates the 2x2 and 3x3 operators with respect to statistical performance. E.g. probabilities of detection vs. false detection for various S/N are compiled.

A discussion of edge detection as pattern classification is also presented, including the application of the Ho-Kashyap algorithm to the problem.

A review of statistical methods is presented, focussing on various methods of hypothesis testing:

Bayes decision rule
Neyman-Pearson criterion
minimax criterion

All evaluations are based on assuming the input to be a perfect step edge plus simple (usually Gaussian) noise. Unfortunately, real data *do not* have step edges and *do* have non-constant areas which are not edges.

An analysis is presented of the effects of Gaussian noise for linear masks.

He uses Pratt's figure of merit to test the various convolution operators. The best performers are the 3 level and Prewitt operators (which are essentially the same). (Pratt's figure of merit is defined as follows. The input is a perfect vertical ramp edge, i.e. a function only of x, having a cross-section of constant-ramp-constant, i.e. a constant part connected by a linear part to another constant part. The variable parameters of the input are the contrast (the difference between the 2 constant values), the slope (of the linear transition ramp), and the standard deviation of additive Gaussian noise. The figure of merit is then defined by a formula based on parameters of the output error. Also, an analogous version is presented for edges at a 45° angle to vertical.)

For convolutions with square support, he analyzes the effects of mask size, center-weighted masks, and local adaptive thresholds.

Abdou proposes 2 new edge operators: 1 and 2 dimensional ramp best fits, resp. The idea for the 1 dimensional case is to fit an ideal 1 dimensional ramp edge to the data for all possible ramp sizes (with discrete end points). Results for each size are given in closed form, but the various sizes must be considered separately to determine the best. The 2 dimensional ramp best fit proceeds in the same way as the 1 dimensional, but he also considers all possible orientations. These he limits to multiples of 45°.

There are several appendices, as follows (with some terse comments).

Appendices
- Analysis of the Hueckel operator (fairly good)
- Orthogonal transformation in edge detection (the beginnings of a DFT method of edge detection)
- The Herskovits algorithm (not a very enlightening discussion)
- Derivations of Eqs. 3.29, 3.31, 3.32 (some statistics)
- Experimental results (pictures) — not very informative, extensive or useful. He only provides binary edge maps of 3 pictures. One can't really see what is happening locally (the pixels are too small to be seen).

Evaluation

To the extent that the local edge ramp hypothesis is valid, the ramp fitting method may work, though it is essentially equivalent to applying various slope masks in various directions. This is a rather inelegant approach since a best fit must be performed for each possible ramp width and angular orientation, with the optimum found by exhaustively comparing the error parameters for all the fits. One advantage over gradient operators and other best fit operators is that the present method can be used to reject regions of smooth shading if the all-ramp condition is rejected as not an edge. However, the Haralick facet model is more general, no more expensive, more elegant, and probably more effective, though probably also inadequate (see review of [Haralick 1980]).

Altes

Altes,R.A., "Spline-like Image Analysis with a Complexity Constraint. Similarities to Human Vision."

The author proposes modelling a (1 dimensional) picture as a finite sum of basis functions which are integrals of delta functions:

$$f(x) = \sum_{n=0}^{N} \sum_{m=0}^{M} f_{nm}\delta^{(-n)}(x - x_m),$$

where $\delta^{(-n)}$ is the nth integral of the unit Dirac delta function, $0 \leq n < \infty$, and the x_m are free knots. Splines can be viewed as such sums with $1 \leq n < \infty$ and smoothness conditions imposed at the knots, hence the paper's title. Including the point spread function, u, of the imaging system yields

$$f(x) = \sum_{n=0}^{N} \sum_{m=0}^{M} f_{nm}u^{(-n)}(x - x_m),$$

where $u^{(-n)}$ is defined analogously to $\delta^{(-n)}$, since $u * \delta^{(-n)} = u^{(-n)}$.

Working in the Fourier domain, he introduces a derived set of normalized basis functions, and shows how to estimate the coefficients in the expansion in the case of a single knot of known position. For multiple free knots, he proposes using techniques from detection and estimation theory, based on a statistical model of knot location. However, the approach is predicated on the use of a matched filter to locate the knots, which appears to be doomed to failure because the basis is not orthogonal. The present writer is not qualified to otherwise evaluate the adequacy of the detection-estimation methods proposed.

The core of the author's method uses filters to estimate coefficients or detect complex patterns. Based on filter complexity considerations, he argues that these filters should all have approximately equal space-bandwidth products. These arguments are related to implementation issues, and for the digital case would be related to cost. One must keep in mind, however, that a major consideration of the work is a theory of human vision. In order to achieve a set of filters with the desired property, he seeks a set of transducer transfer functions to incorporate into the imaging transfer function U. Although it is not stated in the paper, one can think of this as a convolution preprocessor which allows further processing to be done by filters all having the same space-bandwidth product. He uses one particular way of obtaining a constant space-bandwidth product, viz. $V_n(\omega) = \alpha_n V_{n-1}(k\omega)$ for all n with a fixed constant $k > 1$, where α_n is an arbitrary proportionality constant and

$$V_n(\omega) = \frac{U(\omega)/(i\omega)^n}{\|U(\omega)/(i\omega)^n\|},$$

where $\|\cdot\|$ signifies the L^2 norm. Although this is a simple way to get a constant space-bandwidth product, it is not the only way: e.g., a different k could be used for each n. In any case, using this assumption, he arrives at a set of log-normal transducer transfer functions, i.e. functions of the form

$$U(\omega) = A\omega^{\nu} e^{-\rho(\log\omega)^2)}.$$

It is interesting to compare these with the difference of normals suggested by Marr and Hildreth based on similar assumptions.

Some comparisons with human vision are made, notably between line spread functions, which are shown to be similar.

Evaluation

The work is quite provocative. The most interesting features are that it incorporates the transducer transfer function, models images as intrinsically discontinous objects in a coherent way, and uses statistical estimation for detection. Unfortunately, the generalization to 2 dimensions is not easy, a fact which the author does not well appreciate. He proposes 2 routes of "generalization." The more straightforward involves using rasters at a number of angles. Though this is not as satisfying as an intrinsically 2 dimensional approach, it may be a viable way to proceed. Significant problems that would have to be overcome include integrating all the information from the various scan lines (which could be argued to be 99% of the problem to begin with), and accounting for or using a 2 dimensional transducer transfer function. Making a true generalization to 2 dimensions poses the following difficulties. Knots generalize to boundaries. For 1 dimension, the space of knots is 1 dimensional, but for 2 dimensions the space of boundaries is infinite dimensional. The approach of workers in spline theory has been to generalize the intervals between knots to projections from higher dimensional simplices, leading in the 2 dimensional case to piecewise straight boundaries, but this seems to be inadequate for a natural description of the boundary. The 2 dimensional analog of the delta function is a delta function supported by a boundary, but to generalize the integration of the delta function, one requires the analog of the integration of a singular (conservative) force field to obtain a potential, which can be viewed as the integration of a gradient with delta function components, or equivalently the integration of an exact differential, or in modern terminology the integration of an exact distribution-valued 1-form. It is harder to find analogs for the higher order integrals of the delta function, however, although the idea is worth pursuing. Since the main virtue of using the $\delta^{(-n)}$ expansion is the simplicity of convolution with the point spread function, one's efforts probably ought to be directed to preserving that property. What comes to mind immediately is using tensor products of the 1 dimensional $\delta^{(-n)}$ functions, which would be expected to have similar properties, but that would not lead to a simple description of boundaries.

The method of locating the free knots is not very clearly presented, and appears to be based on a possible misconception. In 2 dimensions the problem is of course more difficult because of the complexity of the boundary space. There is no obvious way to solve this problem.

The paper is more biology oriented than computer oriented, so understandably no consideration is given to digital processing issues, the most important of which is the effect of discrete sampling on a periodic grid.

Ballard and Sklansky 1976

Ballard,D.H. and J.Sklansky, "A ladder-structured decision tree for recognizing tumors in chest radiographs," IEEE Trans. Computers, vol. C-25, 5, May 1976, 503-513.

The authors are concerned with finding roughly circular regions of approximately 100 pixels in area.

Summary of processing steps

A thresholded gradient picture is first found using a sequence of processes $\Theta_T \circ \nabla \circ \mathcal{H} \circ \mathcal{L}$, where

$\mathcal{L}$ is the low pass filter operation defined by averaging and then sampling on a coarser grid. (Note that averaging is not strictly low pass, since the filter transfer function is a *sinc.*)

$\mathcal{H}$ is a high pass filtering operation performed in the frequency domain via FFT's, using a filter characteristic attributed to Kruger.

∇ is a digital gradient operator defined in terms of adjacent pixel differences.

Θ_T is a global thresholding operator.

A heuristic search connecting edge pixels, similar to Martelli's technique, is then used to find the lung area, following a Kelly-like "plan."

To locate tumors and nodules, a Hough-like method is used: An accumulator array corresponding to possible circles, indexed by position and radius is incremented by the number of edge pixels with positions and gradient directions consistent with lying on the given circle. An improvement is achieved by using the gradient direction in addition to the magnitude.

Big and small radii are tumor and nodule candidates, resp. The big ones are immediately declared to be tumors, while the candidate nodules are subjected to a 2 stage classifier which looks at features from a detailed nodule boundary finder. The latter is based on growing all optimal edges of length n in a given region until closure is reached, using a Kelly-like "plan."

Evaluation

The accumulator array method seems to be useful for finding some circle-like boundaries. One must always keep in mind 2 questions for such feature detectors: what does it really find, and what will it miss. These questions are best answered either by mathematical proof or application to numerous examples. Unfortunately, neither of these tests is available in the present paper, though probably one cannot fault the authors for not including more examples, since space limitations may have been imposed. In any case, what is being detected is not regions with roughly circular boundaries, but areas having a sufficiently high count of above threshold gradient values (of the right orientation) lying on a circle. This provides some kind of global understanding of the intensity function, which is commendable, but it is not likely to find sharp edges which do not stay near and tangent to some such circle. However, the main use in the paper being reviewed is to guide a more detailed process of boundary finding, and in that context the question becomes whether the feature being found is indicative of a closed boundary in its vicinity. On the one hand, there is little doubt that a roughly circular boundary of adequate contrast, sufficiently defocussed would cause one or a few such circle detectors to fire, allowing the more detailed process to find the precise boundary. On the other hand, the firing of a circle detector is no guarantee that there must be such a boundary: all that is necessary is that the intensity have a steep enough centripetal gradient over a large enough part of a circle, which might happen if the intensity function has a maximum inside the circle.

Beaudet 1978

Beaudet,P.R., "Rotationally invariant image operators," in Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78) (Kyoto, Japan, Nov. 7-10, 1978), 579-583.

The author is interested in finding a least square polynomial approximation to image data. The coefficients of the monomial terms are computed via convolution.

The starting point is to consider the polynomial to be fitted as a truncated Taylor series. The coefficients are found as in a normal least squares problem, but are taken to represent the derivatives in the Taylor expansion. To 1st order, this is the same as fitting a plane and estimating the gradient. The quadratic part is tantamount to finding the classical Hessian.

He considers operators up to 4th order, and operator sizes from 3x3 to 8x8. The only rotationally invariant 1st order operator is the gradient, or rather, more precisely, the squared magnitude of the gradient, $\nabla f \cdot \nabla f$. The 2nd order operators correspond to the linear invariants of the Hessian matrix,

$$H = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix},$$

as well as the scalar valued operators $|H\nabla f|$ and $\nabla f H \nabla f$.

Unfortunately, it appears that the author confuses the Hessian with a matrix representation sometimes called the Weingarten map, which is the differential of the Gauss map. The linear invariants of the Weingarten map are the intrinsic curvatures of the surface: the eigenvalues are the principal curvatures, the trace is the mean curvature, and the determinant is the Gaussian curvature. The author, however attributes these properties to the Hessian. This confusion most likely stems from the fact that the two coincide at any critical point of the function f, and it is possible to rotate the 3 dimensional coordinate system of a surface in $\mathbf{R}^3$ so that any given point is a critical point when the surface is being viewed as the graph of a function from $\mathbf{R}^2 \rightarrow \mathbf{R}$. This is commonly done in expositions of the subject to simplify formulas. However, since we are in a fixed coordinate system, such a simplification is not possible (without, of course, including the rotation matrices). (See, e.g. [do Carmo 1976].) The differential of the Gauss map, when the surface is given as the graph of a function $f : \mathbf{R}^2 \rightarrow \mathbf{R}$, can be written in coordinates x, y as

$$dN = \frac{1}{(1 + f_x^2 + f_y^2)^{3/2}} \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix} \begin{pmatrix} 1 + f_y^2 & -f_x f_y \\ -f_x f_y & 1 + f_x^2 \end{pmatrix},$$

which clearly reduces to the Hessian at a critical point of f.

Beaudet correctly points out that the trace of the Hessian is the Laplacian, but he makes incorrect assertions about the relations between the quantities he derives from the Hessian and various curvatures.

Three 3rd order operators are presented, which are claimed to have significance as line end, curve boundary, and line detectors.

The above terminology and interpretation is ours; he presents these in the more classical language of tensors and coordinates, where his operators are contractions of tensors.

One should note that considerations and techniques very similar to those presented in this paper were described by Prewitt circa 10 years before, though no reference is made to that work.

Evaluation

The experimental results consist in the application of a few of the operators to a single image. Since the notions of line detection and edge detection are very simplistic, there is no effort to use the results of the processing in any way other than to present the magnitude of the operator output. Not surprisingly, this is not very effective. However, more sophisticated processing based on the obtained fit is promising. A potential difficulty may lie in the manner in which the fit is obtained, since polynomial least squares fits tend to produce spurious oscillations.

Despite these shortcomings, the proposal to compute geometrically and analytically significant properties of the image intensity function, using convolutions, is a worthwhile contribution. The thrust, perhaps not made clear by the author, is to derive an understanding of the image intensity function in terms which have precise, well-understood meanings, and which go beyond a few naively chosen parameters. As it happens, the error about intrinsic surface properties may be fortuitous, since it may make more sense to consider the Hessian of the intensity function, rather than its surface geometry independent of coordinate system. There is, after all, a special coordinate system in this situation. It would be interesting to see results of psychophysical studies where the intensity function is changed so that only the Hessian or the Weingarten map, but not both, change.

As presented, this is not a viable edge detection method. However, the idea of local fitting merits further investigation, particularly in regard to deriving differential operators.

Brice and Fennema 1970

Brice, C.R. and C.L.Fennema, "Scene Analysis Using Regions," Artificial Intelligence Group Technical Note 17, Stanford Research Institute, April 1970.

Fennema, C.L. and C.R.Brice, "Scene analysis of pictures using regions", Artificial Intelligence Journal 1, 1970, 205-226.

This is a now-classic work in region growing. Its methods are extremely simple, which *a priori* may not be an indictment, but in this case they are based on a very simplistic image model that was wishful thinking and no one now believes. The approach was motivated purely by heuristics, rather than any theory, and at this level of processing that turns out to be inadequate.

The basic segmentation operation is to partition the image by pixel intensity value. Boundary predicates are based on the completely local nearest neighbor intensity differences.

There are 2 merging heuristics:

Phagocyte heuristic

Merge adjacent regions if the "weak" part of their common boundary is a big enough part of one of their total boundaries. "Weak" and "big enough" are relative to global thresholds.

Weakness heuristic

Merge adjacent regions if the "weak" part of their common boundary is a big enough fraction of it (common boundary). Another global threshold is used for "big enough."

Evaluation

The method presented is much too simplistic. E.g., it will clearly fail if smooth shading leads to 1st differences of the same magnitude as an edge. Noise spikes will always end up as regions. The heuristics are too heuristic — they are not based on any analysis or understanding of real images, beyond a few common-sense notions. Global thresholds are invariably a bad idea: a little observation can persuade one that the same magnitude (of edge parameter, gradient, or whatever) can be significant in one context and meaningless in another.

Chen and Pavlidis

Chen, P.C. and T. Pavlidis, "Image Segmentation as an Estimation Problem," Computer Graphics and Image Processing, February 1980, vol. 12, no. 2, 153-172.

Abstract.

Picture segmentation is expressed as a sequence of decision problems within then framework of a splitrynd-merge algorithm. First regions of a arbitra- initial segmentation are tested for uniformity and if not uniform they are subdivided into smaller regions, or set aside if their size is below a given threshold. Next regions classified as uniform are subject to a cluster analysis to identify similar types which are merged. At this point there exist reliable estimates of the parameters of the random field of each type of region and they are used to classify some of the remaining small regions. Any regions remaining after this step are considered part of a boundary ambiguity zone. The location of the boundary is estimated then by interpolation between the existing uniform regions. Experimental results on artificial pictures are also included.

For simplicity an image is assumed to consist of only two types of regions, each of which is modeled as a collection of Gaussian-distributed random variables of the same mean and variance. The authors claim that the first assumption is easily generalized, and that they are working on more realistic statistical models to replace the second, but offer no details. Their approach is to take each region in some arbitrary initial segmentation and use a standard hypothesis testing technique to decide whether the pixels in the region come from one Gaussian distribution or two, even though there is no a priori information about the parameters of the two distributions. The initial segmentation suggested is a pyramid data structure, where at first there is only one region, which is subdivided into four squares if it is judged to contain more than one distribution; next, each of those four squares is considered separately, and the uniformity hypothesis tested in each, etc. The goal of this initial stage is to form several large regions and to estimate the mean and variance of each; large regions are desirable because large samples give better estimates. Next, these estimates are grouped into two categories characterizing the two different region types; from each category, one region is selected, and its mean and variance used as the final estimate for its region type. Adjacent regions of the same type are merged. At this point, each region not large enough to be part of the parameter estimatation process is tested to decide whether it is a type one or a type two region; the approach is the Neyman-Pearson method, using the mean and variance of the pixel values in the region and the means and variances of the two region types in the likelihood ratios.

Because the confidence of a given region's assignment to region type one or two decreases with the size of the region, it is likely that small regions near a boundary between the two region types will be misclassified. Instead it may be desirable to leave regions below a cutoff size unclassified and to consider them a part of a border ambiguity zone. In fact the authors show that it is not possible to determine the boundary exactly, so that the ambiguity is intrinsic to the problem, not to a given method of solution. They suggest, however, a heuristic method of estimating a boundary using the ambiguous region as follows: construct a center line through the ambiguous region halfway between its left and right boundaries; for each point on the central line, calculate the absolute value of the difference between the average of the pixels on the left side along the perpindicular to the central line and that on the right side; the boundary is estimated by linking all points on the central line representing local maxima of differences to either side. (There is also some use of the local minima which I don't quite understand.)

Experimental results on synthetic images with the statistical properties n which the algorithm is based are described. The boundaries in the images are found successfully even where visual discrimination is very hard.

What is useful in this paper is the competent and thoughtful application of statistical decision theory and estimation theory to a simple two-dimensional domain: it is possible to learn quite a bit about how to use these tools from a careful study of the methods presented. Of less obvious value is its relevance to natural images. Natural images contain many region types; each region is more a smoothly undulating surface on which is superimposed noise, and only the noise lends itself to simple statistical characterization; use of spatial information within regions is thus critical; boundaries between regions are blurred. These failings have much in common with those of other statistical approaches: an inadequate image model, and ineffectual use of spatial information.

Cooper et al. 1980

Cooper, D.B., H. Elliott, F. Cohen, L. Reiss, and P. Symosek, "Stochastic Boundary Estimation and Object Recognition," Computer Graphics and Image Processing, vol.12, 1980, p. 326.

Abstract.

The authors present a maximum likelihood approach for finding the boundary of a blob of constant grey level on a background of constant but different grey level, with gaussian noise superimposed on both. An application to object recognition is discussed.

An image is modeled as object of constant grey level with a highly varying boundary, surrounded by a constant grey level background. A white Gaussian noise field is superimposed. The boundary is modeled as a Kth order Markov process as follows: an edge element is a boundary between two adjacent pixels, and an object boundary is a sequence of such non-self-intersecting elements. Given a list of K such boundary elements, the K + 1st element can be one of three edge elements; the probabilities of these three possible choices, by the Kth order Markov property, depend only on the K preceding edge elements. Thus, if we ignore for the moment the fact the a boundary surrounds an object, the probability of a given boundary, given the probability of its starting location and that of its first K elements, can be calculated if all the transition probabilities are known. Moreover, it is possible to select a set of transition probabilities to "favor" long straight boundaries over highly varying ones by setting the probabilities corresponding to straighter segments higher than the others; the authors do so in their experimentation with synthetic images.

Given the grey levels of object and background, the variance of the noise (mean 0 is assumed), and the boundary-related probabilities described above, the joint likelihood of any hypothesized boundary and the entire picture function can be computed. The "ripple filter" described in this article is based on comparing this likelihood to that associated with a slight perturbation of the boundary. The algorithm is basically to choose a point on an initial boundary found by any method and, in the simplest case of a long, straight segment, compute likelihoods associated with three alternatives: first, that the boundary remains where it is; second, that the two pixels separated by the edge element are both object (i.e. boundary should be moved to include them both); and third, that the two pixels are both background (i.e. boundary should be moved to exclude them both). At a corner the situation is somewhat more complicated and requires the evaluation of four likelihoods, but the idea is the same: the most likely case, according to the measure defined, is taken as true and the boundary is modified accordingly. Then, the next edge element along the boundary is tested and modified in the same way; then the next, etc. Many tours may be required before convergence, with the boundary tending to grow within the object and shrink in the background. The authors do not analyze the conditions necessary for convergence.

Because the evaluation of likelihood depends on computations with a number of pixels that is a function of the order of the Markov process describing the boundary, the ripple filter can be quite slow. The authors discuss various ways to speed up the computation of likelihood, including the use of lookup tables, alternative state structures, and varying resolutions.

The authors describe a Sequential Boundary Finder as an alternative to the ripple filter for maximizing the likelihood measure. Computations are restricted to windows for efficiency, and possible extensions of the boundary are considered using a search procedure based on the A* algorithm. The idea is to make local decisions in an optimal way.

Next an alternative boundary model is proposed. The one described above follows from making change in angle a function of arc length. Also mentioned is a model based on a discrete-time linear dynamical system driven by sequences of independent, identically distributed random variables.

Several pages are devoted to boundary error estimation, but in a context slightly different from that described above. The boundary is specified by a sequence of horizontal edge elements: i.e., in the x-y plane, the boundary is defined by a sequence of y-values, each constant in the interval between equally-spaced x-values, and vertical segments corresponding to the jumps at the x-values. Below the boundary is the background, and above it is the object. The model for pixel values in the object and the background is the same as before. By treating the y-values in each strip (i.e. an interval between two x-values) as independent random variables, and using some standard signal detection theory, the authors show that the estimate for a single y-value has mean equal to the true y-value and variance bounded by a function they derive. Next, by modeling these estimated y-values as a stochastic process with additive white Gaussian noise, where the stochastic process is the true y-values and the difference between the true-values and their estimates is the noise, they use signal estimation theory to derive the minimal achievable error variance. The true y-values are modeled as the sum of a known mean value function and a zero-mean stationary stochastic process generated by driving a constant coefficient linear difference equation with white Gaussian noise. Throughout the boundary model is one-dimensional.

A possible application of this model to object recognition is discussed. The image model of before, an object on a background with superimposed Gaussian noise, still holds, but now the object can belong to one of many classes, each of which may have a different expected grey level, boundary model, and a priori probability of occurrence. An approximation to the Bayes object classifier is derived for most cases of practical interest. The application of the boundary error estimation model described above to the analysis of recognition error is discussed briefly.

Criticism

The structure of their theory is appealing: it is comprehensive, cohesive, systematic, and quite rigorous (as far as I can tell). There is very little of the flavor of ad hoc methods, of techniques proposed and implemented because they seem to work well, not because they can be justified from first principles. But it is hard to see how their theory could be used for edge detection in natural images. Bits and pieces seem relevant, such as the maximum likelihood method for perturbing a boundary to find the most likely boundary, but the statistical models proposed seem far too simple; and more realistic ones could make some of the computations intractable, particular the more global ones.

Davis 1973

Davis,L., "A Survey of Edge Detection Techniques", TR-273, Univ of Md, Computer Science Center, 1973.

The author presents some discussions of prior edge detection techniques:

Parallel edge detection
- Herskovits and Binford
- linear vs. nonlinear operators (nonlinear mainly Rosenfeld)
- texture edges
- Griffith
- Hueckel
- Chow

Sequential edge detection
- Martelli (simple, not the more general)
- Montanari

"Guided" (top-down) edge detection
- Kelly
- Harlow
- Shirai

He discusses and criticizes what was done very tersely. There are no particularly deep or sophisticated analyses; nevertheless this work provides a useful first tour or refresher of some of the more significant work in the field. One can detect a subtle and not surprising bias toward Rosenfeldism.

Dreschler and Nagel 1981

Dreschler,L. and H.-H. Nagel, "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI-81), Aug 1981, Vancouver.

Dreschler,L. and H.-H. Nagel, "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," Report IfI-HH-M-90/81, Fachbereich Informatik, Universität Hamburg.

The authors are primarily interested in tracking objects in a sequence of successive static frames. They seek point features which are expected to be stable from frame to frame, settling on extremal points of the Gaussian curvature of the intensity function. The computation of the curvature is performed via "principal curvatures" using the operators of Beaudet (which in fact compute something other than principal curvatures: see review of [Beaudet 1978]).

The authors are motivated by seeking local extrema of Gaussian curvature. However, they found that such extrema occur at knees of edges (cliffs in the intensity function) in an unstable manner, as a consequence of local noise and small. Therefore, a more involved predicate is used. Viz., pairs of nearby points are found which are a maximum and a minimum of Gaussian curvature. Along the line joining these points, that point having the steepest slope of intensity (i.e. directional derivative) is selected as the feature point, subject to the following 2 criteria. First, it is asserted that exactly 1 principal curvature must change sign along the line in question (this is true only if the extrema of Gaussian curvature are of opposite sign, which is implicitly assumed), hence it is required that the principal direction corresponding to the principal curvature which is changing sign be roughly parallel to the line in question. This assumes that the extrema of the Gaussian curvature should be joined by principal curves, a proposition whose truth is by no means self-evident. Secondly, the intensity value at the maximum must be greater than that at the minimum. This is for the case that the high intensity area is convex at the corner. Since the reverse case obtains by turning the surface upside down, which does not change the Gaussian curvature anywhere, the opposite condition must be true when the low intensity area is convex, so without other information about the context of the extrema, this seems to be a vacuous condition. Also, an *ad hoc* maximum separation of 4 pixels is required for pairs of extrema to be linked. Obviously, this is a requirement that the corner be quite sharp at the resolution of the image.

Both 5x5 and 3x3 operators are used: the 5x5 for good noise behavior, and the 3x3 for better resolution in places selected by the 5x5. The operators used are the ones presented in [Beaudet 1978]. Consequently, the present authors are victims of an incorrect definition of Gaussian curvature (see review of [Beaudet 1978]) and principal directions. However, it is extrema of Ga ıssian curvature which are of interest. The relation between these and what is actually (erroneously) used is algebraically complicated, and we do not attempt to analyze it, but these paramaters may be just as meaningful for images as the geometric ones. Furthermore, there is already a heuristic element to locating the points of interest. Therefore, it doesn't seem likely that the use of the correct values of the Gaussian would change the performance significantly. To get a better understanding of the situation, one should in fact analyze the behavior of these parameters in the light of what is known about the image irradiance equation.

Experimental results

The results displayed seem to be fairly good. Of course, there are a number of other elements of the system we are not considering here, e.g. the method of tracking, so that it is difficult to say how reliably the features selected represented intrinsic features of objects or even of the intensity function.

Evaluation

The present work is best regarded as a corner detector. As such, it is not adequate for performing segmentation. As far as its usefulness for matching images is concerned, one would have to analyze to what degree extrema of Gaussian curvature are intrinsic features of the object geometry, rather than the intensity surface geometry. There are 2 components to such a study: the effects of perspective transformation, and the effects of photometric laws. An initial approach could consider these components separately, i.e. constant light with moving observer, and fixed observer with moving light source. Since the features used are piecewise smooth functions of the parameters of motion and lighting, one can expect that they will trace out piecewise smooth curves as those parameters are varied; and hence they can be tracked. Whether they are good things to track is another question. Consider the extreme case of a moving flat mirror, moving in its own plane, and reflecting a light source. This isn't a completely ridiculous case, since it is a limiting case of what can happen with specularity, which in turn is a matter of degree for the reflectance function. The point to note is that the feature associated with the specularity will behave as a function of the location of the light source rather than as a function of the motion of the object reflecting it. The moral is that the behavior of a feature can be highly decoupled from that of the object whose surface creates it.

The relevance to image segmentation is this. Principal curvatures, principal directions, and principal curves are useful features of the image intensity function. They define a local geometry, and notably a local orthogonal coordinate system which is a natural coordinate system in the vicinity of edges. Predicates based on observation of the behavior of principal curves seem good candidates for edge detection and hence segmentation. This work shows at least that such features have some tability in the presence of noise and deformation.

Duda and Hart 1973

Duda, R.O. and P.E.Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

Abstract.

Only pp. 338-339 are reviewed here. A method of curve segmentation called iterative endpoint fitting is described.

The iterative endpoint fit algorithm consists of the following (this is more of a paraphrase than a critical review). A line is fit to an initial ordered set of points by connecting the endpoints. The distance from each point to the line is computed, and if no point is further away that some threshold, the process is finished. Otherwise, the original line is broken in two by drawing lines connecting each endpoint with the point farthest from the original line. Then the process is applied to each of the two new lines. They mention two drawbacks to this process. The first is that a final segment may not be a very good fit to the points in its immediate vicinity, but fortunately a post-fit process to adjust line segment positions can remedy this problem. The second drawback is that one wild point can dramatically change the result of the process. A preliminary smoothing step reduces the impact of wild points but causes other problems. They suggest that this algorithm is best suited to largely noise-free data.

Griffith 1973

Griffith,A.K., "Mathematical Models for Automatic Line Detection," Journal of the ACM, vol.20, no.1, January 1973, p. 62.

Abstract.

A particular decision-theoretic approach to the problem of detecting straight edges and lines in pictures is discussed. A model is proposed of the appearance of scenes consisting of prismatic solids, taking into account blurring, noise, and smooth variations in intensity over faces. A suboptimal statistical decision procedure is developed for the identification of a line within a narrow band in the field of view, given an array of intensity values from within the band. The performance of this procedure is illustrated and discussed.

Edge and line detection is based on rigorously computed conditional probability that, given image data, one can evaluate the probability that the data contain a lien or an edge. This standard decision theory framework enables parameterized control of the decision process by guaranteeing, for example, various combinations of minimum rates for false positives and false negatives. The underlying models of lines and edges which enable this calculation are well-specified, but at a large cost in generality: the scene consists of prismatic (plane-faced) solids uniformly white in reflectance; there are no shadows; image intensity varies smoothly over each face of the objects, and in the background; and all edges are either steps or lines (i.e. an extended impulse) with consistent amplitude and type (i.e. edge or line) along an edge or line. Moreover, the detection process takes as input a rectangular subregion of the image, constant in size and shape, and decides only whether or not there is a single line or edge at a vertical orientation exactly centered in the rectangular region. No provision is made for distinguishing among those cases for which the vertical, centered line or edge is not present, for example, regions containing an edge or line at a random location and orientation, or those containing more than one possibly intersecting line or edge.

Simple mathematical models of three classes of input data, regions containing lines, those containing edges, and homogeneous regions, are developed; basically, an edge is modeled as a plane with a step in it, an line as a plane with a ridge of impulses, and a homogeneous region as a plane. Convolution with a point spread function and superposition of Gaussian noise transform these models of scene events, into models of image events. Simplifying assumptions here include that a single edge profile and a single line profile differ from the universe of possible edge and line profiles in the absence of noise by at most a multiplicative constant. Algebraic sleight-of-hand of the conditional probabilities of a set of input data given that the region falls into one of the idealized classes eventually yields the final decision function: the conditional probability that the region of interest contains a line or an edge, given the input data. Subexpressions of the decision function are shown to have intuitive significance.

Empirical investigations were undertaken to verify the underlying assumptions and to evaluate the performance of the decision function. Verified assumptions include: the noise in the image device is Gaussian; edge type profiles differ only in amplitude, not in width or skewness. Rather than two basic edge types, however, six (to a reasonable approximation) were found. Fortunately the decision function can easily be extended to include six edge types. The assumption that intensity scans normal to an edge were consistent along the edge in amplitude and type turns out to be more true of intense than of subtle edges.

The size of the sampling domain, i.e. the input to the decision function, is discussed: the two heuristically developed guidelines are: the width of the field should be just enough to contain the entire edge profile to avoid the addition of more edge profile types to discriminate among the different "tails" in each of the six classes, and the number of samples in a line should guarantee that a scan across a line include at least one value close enough to the peak to have 90 percent of its value.

A localized version of the decision function requiring only one scan line is derived and its performance compared to that of the Roberts' cross operator [Roberts 1963]; the decision function turns out to be twice as sensitive as Roberts' operator in terms of the faintest edge detectable at given noise and confidence levels. Note that within one scan line edges can only have one orientation, so the fact that the decision function only detects vertical edges is not a problem.

The most obvious weakness of Griffith's approach is that its restriction on orientation and location of edges it is designed to detect make it inapplicable as it now stands to real world images (except for the single scan-line version). Other simplifying assumptions limit the number of basic edge types, but it is unknown how many there are in the real world, so that the computational complexity as a function of the number of edge types becomes an issue. Understanding how to calculate an edge-detecting decision function is useful, but the domain here is so simple that the workability of the method in more realistic domains is unclear.

Haralick 1980

Haralick, Robert M.; "Edge and Region Analysis for Digital Image Data," Computer Graphics and Image Processing, vol.12, no.1, January 1980, 60-73.

The view taken here is that edges and regions can be viewed as places where there are large or small differences, resp., in some parameters. In this light, the old method, i.e. looking for perfect step edges, amounts to fitting a piecewise constant function to the image intensity. The new method which the author puts forth, is to do a piecewise *linear* fit, i.e. to fit planes (or *facets*). The work is purely theoretical in that real images are not considered.

The central feature of the analysis is to perform a least squares fit of a plane to the data. The author provides a nice analysis of noise for this problem. The critical question is whether 2 planar patches are actually part of the same plane: the edge null hypothesis. To resolve this question, he uses the F-test on a χ^2 distribution derived from the error.

More specifically, the way this is used is as follows. Each point p of the picture is assigned a neighborhood $p \mapsto U_p$ which is the one supporting the best fit among all U_p^i containing p. I.e., of all neighborhoods U_p^i such that $p \in U_p^i$, let U_p be such that $\epsilon(U_p)$ is minimum.

Edge and region detection are then based on an F-test of the parameters associated with the optimal neighborhoods for adjacent pixels, followed by thinning. Even neglecting the piecewise planarity assumption, this adjacent-F-test is probably too simple minded.

The technique can be summarized as follows:

edge detection method:
each pixel has a best-fit neighborhood with parameters of fit.
edgeness=F statistic that adjacent pixels' fits come from same plane
compute for vertical, horizontal adjacencies for vertical, horizontal edges
find maxima by non-maximum suppression

region growing method:
group adjacent pixels if same best fit neighborhood plane hypothesis cannot be rejected

The hypothesis testing is based on the relation between parameter differences and errors of fit. If the local is relatively poorer, greater parameter differences are tolerated for region merging. In this sense, the region merging is adaptive. However, no analysis is presented describing how this method would behave for large regions or long edges. Also, no attention is given to the problem of determining whether local edges are part of a larger edge.

The author includes quick but nice review of related literature. For example, he shows that Robert's cross operator [Roberts 1963] is the magnitude of the gradient of a linear fit (although this is all but explicitly stated in [Prewitt 1970]).

Unfortunately, the paper includes no experimental results or consideration of real images.

Evaluation

The idea of fitting regions and looking at the parameters is a good one. Statistical analysis is good, too. However, the piecewise planar hypothesis is not sophisticated enough. On the other hand, the statistics becomes more complicated for more complicated fits. In the form proposed, this method is not likely to be noticeably better than other local methods. The extended edge and region part is rather ad hoc — not based on a sound analysis. This paper can be recommended as a good introduction to the use of statistics and fitting, despite some ambiguities.

Haralick 1981

Haralick, Robert M.; "The Digital Edge," Proc. IEEE Conf. Pattern Recognition and Image Processing, August 1981, 285-291

The essential feature of the technique proposed by the author is fitting the image intensity function by a polynomial.

He first defines edges as discontinuities in brightness value or its "derivative." But then he notes that for this to make sense, the discrete picture must be thought of as samples of a function on a continuum. He then proceeds by taking the tone of making a derivation of requirements for an edge finder, but the requirements put forth are inadequately supported, and far from certain. E.g., he requires the assumption that the derivatives of the underlying function are uniformly bounded except at discontinuities (so that high estimated values can be attributed to discontinuities).

He does polynomial approximation using discrete orthogonal polynomials. "Discrete orthogonal" means orthogonal with respect to the "inner product"

$$(f,g) = \sum_{p \in P} f(p)g(p)$$

where P is some finite set of points, though this is not explicitly stated. It is not a true inner product because it can happen that $(f,f) = 0$ with $f \neq 0$. (see e.g., [DeBoor 1978]). Regrettably, he provides no references: there is, after all, a rather large literature pertaining to fitting polynomials.

In the context of fitting functions, he defines an edge to be a place where the "direction isotropic magnitudes" of the 1st or 2nd partials of the fitted function exceed some threshold. This is slightly naive, but on the right track, viz. looking at the parameters of a fitted function.

He imposes 1 dimensional symmetry on the index sets of the polynomials, i.e. the points at which they are defined must be symmetric about the origin. For 2 dimensional basis functions, he uses the tensor product of his 1 dimensional set. He then shows how to fit by the usual method of projection onto the orthonormal basis. A further section is devoted to showing that $D_x^2 + D_y^2$ and $D_{xx}^2 + D_{yy}^2$ are rotationally invariant differential operators.

Evaluation

The idea of fitting a function to the intensity data as a first step in edge finding is good, although the definition of "edge" is somewhat simple-minded. E.g., the 1st derivative criterion will result in edges being found in regions of smooth shading. Unfortunately, the paper itself does not say much, though one presumes that a reasonable amount of thought has been given to the fitting problem. E.g., polynomial least squares fits (which are being proposed) are notorious for being badly behaved — they tend to have extra wiggles. One would expect that such functions would not be very good ones to use if one wanted to look at derivatives. No mention is made of his previous idea of looking at discontinuities of *parameters* of fit between adjacent regions. Nevertheless, some kind of fitting process seems to be in order to use global information for local features (in this case the global fit yields the local derivative). The noise performance issue is promised for the future; this is an important problem — without even considering optimality, one still must worry about the stability of the fit. I.e., the final estimate should be C^1 with respect to the data, or, if not, the discontinuities should be understood.

This should be regarded as "throwing out an idea" rather than as a completed product. The specific edge detector proposed is not likely to have outstanding performance, but other operators arising from the same idea are a promising area of research.

Hough

Hough,P.V.C.; "Method and Means for recognizing complex patterns," U.S.Patent 3,069,654, December 18, 1962.

Duda,R.O. and P.E.Hart, "A generalized Hough transformation for detecting lines in pictures," SRI AI Group Tech Note 36, 1971.

Duda,R.O. and P.E.Hart, "Use of the Hough transformation to detect lines and curves in pictures," Comm. ACM 15, no.1, 1972, 11-15.

Duda,R.O. and P.E.Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

The Hough technique offers a solution to the problem of finding global straight lines, or more exactly, finding global sets of nearly collinear feature points. In the present context, "global" means over the entire image, though other workers have used the same idea for subregions.

Basic idea:

Consider **L**, the set of all lines in the plane, as a topological space. Duda and Hart use the so-called normal parametrization for **L**, where each line is specified by the pair (θ, ρ), representing the orientation and distance from the origin of the line. This parametrization is borrowed from integral geometry, where it is used in the solution of the Buffon's needle problem. It derives its utility from providing a translation invariant measure for the space, so that probabilities behave in desired ways. ([Santalo 1976] is an excellent source for information about integral geometry, and should be of interest to vision researchers.) Hough, on the other hand, used the slope-intercept parametrization familiar from analytic geometry, but which is fraught with difficulties for this situation. Incidentally **L** is a non-trivial space: for $\rho > 0$, every value $0 \leq \theta < 2\pi$, defines a different line. But when $\rho = 0$, i.e. for lines through the origin, (θ, ρ) defines the same line as $(\theta + \pi, \rho)$. Thus, **L** is homeomorphic to an semi-infinite circular cylinder with the bounded end terminated so that antipodal points on the cross-section circle are identified, which in turn is homeomorphic to a punctured disk with antipodal points on its periphery identified. This is also the same thing as an infinite Möbius strip, formed by taking a doubly infinite strip and gluing it together with a half-twist.

The basic insight Hough used is this. For each point p in the plane, there is some curve $\gamma(p) \subset \mathbf{L}$ which corresponds to all the lines through p. For each p of interest in the picture, accumulate weight for $\gamma(p)$ in **L**. Then lines in the picture will be places in **L** with high accumulated values. (One can think of this as defining an weight accumulation function $h : \mathbf{L} \rightarrow \mathbf{R}$ by $h = \sum \chi_{\gamma(p_i)}$ where $\chi_{\gamma(p_i)}$ is the characteristic function of $\gamma(p_i)$.

As Duda and Hart point out, the method provides a savings because of quantization of **L**. The finer the quantization, the less the savings.

Evaluation

The Hough method is not adaptable beyond very limited spaces of curves because storage requirements grow exponentially with the number of parameters characterizing the features, i.e. with the dimension of the space of curves.

Since the method is totally global, undesired features can come into play, i.e. the noise level is high due to many chance contributions throughout the image. However, to combat this problem, one can design localized variations, at the price of requiring a method to patch the local results together.

Possible generalizations include the detection of curves with more parameters, weighted accumulation (based on confidence or significance of the data points), the inclusion of noise considerations, and localization.

The success of the Hough method is very dependent on selecting the initial points of interest, i.e. on the local feature operator. On the other hand, a good way of doing this might compensate for large parameter spaces.

Hsu, Mundy and Beaudet 1978

Hsu,S., J.L.Mundy, P.R.Beaudet; Web Representation of Image Data,in Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78) (Kyoto, Japan, Nov. 7-10, 1978), 579-583.

The authors are interested in using a local quadratic fit to detect image features. A quadratic polynomial is least squares fit to the image data on (presumably uniform) local neighborhoods. The polynomial is regarded as a Taylor series, and the coefficients are interpreted as partial derivatives (see [Beaudet 1978]). The "principle [sic] axes" are identified, and a mesh is constructed over the image by extending straight lines along these special directions until some error threshold is reached, resulting in a new mesh node and repetition of the process. The implementation is based on starting from seed nodes, with special rules for the image periphery, propagating down and right, and merging of nearby nodes. Some nodes of the resulting mesh are labelled according to the "curvatures" and an extremum predicate. Global paths through the mesh are then sought by the use of production rules based on the local labelling to follow arcs. It is not entirely clear how this process works; apparently some kind of relaxation is involved.

Experimental results

Partial results are shown for 1 real and 2 synthetic images of ca. 128x128 resolution. Feature finding is only shown for two of these, where a purported ridge is found in a synthetic normal saddle, and some ridges are found in a real picture of scratches. The performance on the real picture is quite poor, although it is hard to isolate the reason. Probably it is a consequence either of the extreme coarseness and irregularity of the mesh, or the localness and ingenuousness of the production rules.

Evaluation

See the review of [Beaudet 1978] for remarks about fitting and differential operators. The same misconception is present here as in [Beaudet 1978] regarding use of the Hessian to define principal curvatures and intrinsic surface properties, rather than the correct expression for the differential of the Gauss map. Consequently, the "principal axes" and "curvatures" the authors find correspond to the conventional usage of those terms only at stationary points of the image intensity function. However (see review of [Beaudet 1978]), these objects may actually be more meaningful for image analysis than the geometric invariants.

The construction of the mesh is a good idea insofar as a coordinate system based on principal directions is found. However, the mesh is far too coarse, and the method of its construction leads to a topology which may not have much to do with the underlying structure. The authors apparently wanted a graph structure to propagate their production rules on, but unless they have bugs, what they got was more or less a mess. The production rule technique is not very well explained, hence difficult to evaluate, but the impression one gets is that it is somewhat inflexible, e.g. putting limits on rotation of principal direction. It is not clear, e.g. how the production system performs a function separate from the mesh generation itself, where error criteria are also imposed. It may be that using a finer mesh would provide much improved results.

A second problem is that no analysis is given regarding noise behavior. A big question is the behavior of the mesh generation in the presence of noise.

Hueckel

Hueckel,M.H., "An Operator which Locates Edges in Digital Pictures," Stanford Computer Science Dept. Memo AIM-105, Oct. 1969.

Hueckel,M.; "An Operator which Locates Edges in Digital Pictures," JACM, Vol.18, No.1, January 1971, 113-125. Erratum in 21, 1974,350.

Hueckel,M.; "A Local Visual Operator Which Recognizes Edges and Lines," JACM, Vol.20, No.4, October 1973, 634-647.

[Abdou 1978] presents a detailed analysis, to which we direct the reader rather than repeat the same points.

Hueckel, in these papers, seems to be trying to appear more mathematically sophisticated than might be warranted. An example is his statement that "The set of all continuous functions over [the closed unit disk] is a Hilbert space." Since a Hilbert space is defined to be a complete normed inner product space, the statement is false because the space in question is complete in the sup norm, where there is no inner product, but not complete in the inner product space L^2, which is the one Hueckel is using. This brings other mathematical claims which are not proven under suspicion, e.g. the statement that the basis functions are the unique solutions of some unspecified set of "functional equations."

The method involves finding the parameters of the best fitting ideal step edge in a disc-like region of 32 to 137 pixels. The fitting is done in the spirit of the Rayleigh-Ritz method of finding approximate solutions to variational problems (see, e.g. [Morse 1953]). Using a fixed orthonormal basis for the function space of interest, and a fixed truncation of the orthonormal basis, he finds parameters to minimize the L^2 distance between the projections of data and ideal edge in the finite dimensional space spanned by the truncated series. I.e., let ψ_i, $i = 1, \cdots, \infty$ be an orthonormal basis for L^2. Let $f : \mathbf{R}^2 \rightarrow \mathbf{R}$ be the picture (data). Let $E_{\theta,p}$ be an ideal step edge of orientation θ centered at $p \in \mathbf{R}^2$. Consider the space S_k spanned by the first k basis vectors, and let π_k be the orthogonal projection onto that space. Then the idea is to minimize $d(\pi_k f, \pi_k E_{\theta,p})$ with respect to θ, p. Since the basis is orthonormal, this can be done componentwise. This is computationally efficient because the series is truncated at a point which allows a closed form solution for the least squares problem. The line paper uses essentially the same method, with additional parameters to allow for an ideal step edge-line, i.e. a sum of 2 parallel ideal step edges. The method can equivalently be thought of as fitting the best function from the fixed subspace S_k to the data and finding the best edge fit to the function. (This is a consequence of orthogonalities of various subspaces).

The orthonormal expansion used consists of polynomials in x, y with a uniform radial weighting function $\sqrt{1 - x^2 - y^2}$. For the edge (old) operator, 8 polynomials up to degree 3 are used, while the edge-line (new) uses 9 polynomials up to degree 4 (neither set spans the space of all polynomials up to their maximum degree). What, if any, classical set of orthogonal polynomials these correspond to is not stated and not immediately evident, since the definition of the basis functions is presented in a complex way. The orthogonal functions are related to a Fourier-Bessel basis, since $x = r\cos\theta$, $y = r\sin\theta$, and the r polynomials can be thought of as approximations to the Bessel functions one obtains for a radial Fourier transform. It is not stated how the basis functions were derived, however.

The edge/no-edge decision is based on the "angle" between the projections of the data and the best fit edge in the truncated space S_k. I.e., he thresholds on the value of

$$\frac{(\pi_k f, \pi_k E_{\theta,p})}{|\pi_k f| \cdot |\pi_k E_{\theta,p}|}.$$

This suffers from the common problem that little analysis is devoted to the possible picture functions $\pi_k^{-1}(\pi_k E_{\theta,p})$, which are going to look like edges to this operator. In particular, the average gradient plays a large role, and the decision criterion therefore tends to respond to areas with large average gradients over the support.

Evaluation

The main contribution here is to approach the best edge fit problem in a tractable subspace, thereby transforming an essentially combinatorial problem into an analytic one. The particular implementation of that idea, however, suffers numerous shortcomings.

Several criticisms have appeared in the literature.

Abdou, in his thesis [Abdou 1978], argues that the truncation of the orthogonal series introduces excessive error, especially for thin lines, and that unjustified assumptions are made in the optimization procedure. Shaw [Shaw 1977, Shaw 1979] makes a similar criticism of the optimization. [Davis 1973] complains that no attempt is made to relate performance to the image noise process.

Experience using the operator shows that regions of smooth shading result in multiple firings, while regions busier than the size of the operator have missed edges and poor parameter values. These failures are a consequence of using a poor model for the underlying image intensity function. The edge and edge-line models are unrealistic, especially for the support area of the operator. The difficulty can be traced to the fact that in the spaces considered, ideal edges and linear functions are not mutually orthogonal.

Unfortunately, no analysis exists, either here or elsewhere, of the error one incurs by using such simplistic models.

Kanade 1978

Kanade, T., "Region Segmentation: Signal vs. Semantics," in Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78) (Kyoto, Japan, Nov. 7-10, 1978), 579-583.

A survey of image segmentation is presented, based on the paradigm Image → Picture Domain Cues → Scene Domain Cues → Model → Instantiated Model → View Sketch → Image

..., which may be iterated. The distinction is made among the categories of signal, physical, and semantic knowledge.

A large number of works are briefly surveyed, and categorized according to how they fit into the above paradigm. For example, many methods use only signal level knowledge, and hence, in this paradigm, can provide at most a segmentation based on picture domain cues.

Evaluation

The paradigm presented can be more conventionally summarized as saying that one's goal must be to infer the 3 dimensional structure of a scene in order to model the scene and understand the image. Furthermore, one must use physical knowledge, e.g. imaging physics and geometry, to make this inference. This is hardly new or controversial. What is debatable, however, is the distinction which is made between picture and scene domain cues. The main orientation of the paper is toward region growing and splitting methods, using fairly primitive "signal level knowledge," e.g. histograms of the image gray values. For these types of systems, the image-picture-scene-model division is clear and seems natural. But for "image understanding" in general (which the author is addressing), such a glib description does not seem justified, and no arguments are presented to persuade the reader, though in the author's defense it must be said that there were severe space limitations for a fairly broad article.

It seems reasonable that the first step in image understanding might well be to compute a description of the image data in a more useful representation, or set of representations, than is provided by the standard one, i.e. the set of pixel values. Kanade notes, in fact, that [Pavlidis 1972] defines segmentation as a process for describing the image features themselves. From this point of view, "picture cues" are features of this re-representation. (Kanade takes a more restrictive and confused view; he defines "picture cues" by the examples line segments, homogeneous regions, and intensity gradient. The last of these is properly a property of the image, but it can be argued that the first two generally cannot be extracted reliably without using knowledge about 3 dimensional structures, and that is tantamount to making inferences about the "scene domain," although admittedly historically such inferences are implicit.) But it is not so obvious that there must be a trichotomy picture-scene-model. First, the new image representation is chosen based on physical knowledge — the knowledge that determines for what it is important to look. Whatever features are focused on in analyzing the new image representation are likely to be interpretable as features in the scene domain only in conjunction with fitting them into a model. For example, the interpretation of a narrow gradient shaded region may depend on its connection to other regions and on some set of hypotheses about other regions in the vicinity. This might even be on the level of deciding whether the region is an object limb, a surface, a highlight, or even whether it should be regarded as a separate region at all. One can readily envision a Waltz or Zucker type relaxation process occurring using the semantic relations of a model to interpret part of an image representation as a scene feature. In the shape-from-shading paradigm one is hard put to identify any stage identifiable as "picture domain cues."

In summary, the paradigm presented is a useful one for discussing extant image understanding systems, and is particularly clear for those based on the more rudimentary image characteristics. One must be careful, though, not to be misled into a dogmatic adherence to the paradigm presented, since it seems likely, perhaps necessary, that it is inadequate as a description of the type of system required to do successful image understanding in unrestricted environments. The survey is readily accessible as well as concise; it is recommended as a good entry into a fair portion of the segmentation literature.

Kirsch 1971

Kirsch,R.A., "Computer Determination of the Constituent Structure of Biological Images," Computers and Biomedical Research, Vol.4, No.3, 1971, 315-328.

The author indicates that he is Interested in image processing as deriving data structures from image data.

He differentiates between *"well-defined objects"* and *"aggregates,"* which is essentially the difference between smoothly shaded objects with smooth boundaries, and textured "objects" with texture boundaries. His examples are, among others, cells as well-defined, but tissues as aggregates.

The goal is to find boundaries for both types of objects, and the approach is via a local contrast function which is based on the use of the convolution masks

$$\begin{array}{rrr} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{array} \qquad \begin{array}{rrr} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{array}$$

and their 90° rotations. The local contrast function C is then defined as the local *max* over all masks of the absolute values from the convolutions. He defines a *blob of heterogeneity* K as (our equivalent definition) a connected region R such that $C|_{intR} \leq K$ and $C|_{\partial R} > K$: basically a low contrast region with a high contrast boundary, with "low," "high" defined by the threshold K.

The data structure he derives is based on the observation that varying the threshold induces a partial order on the regions by inclusion, which is of course a functorial consequence of the natural ordering on the thresholds. This partial order he represents as a tree, and as a reduced tree showing only when regions coalesce.

Evaluation

This is better than just intensity thresholding, but suffers many of the same drawbacks. Although he keeps track of what may happen for all threshold values, the thresholds are still *global* thresholds, although one could generalize slightly and use thresholds global only to a region. Now although one might expect boundary contrast to be less variable over some region than simple intensity, it's easy to imagine, e.g., a weak spot in a boundary such that lowering the threshold to include the weak spot introduces enough boundary points to disconnect the region.

Because only the *max* of the directional contrasts is used, important geometric information is discarded in finding the boundaries. This is apt to lead to errors for uniform regions, since noise cannot be rejected on the basis of direction to other boundary points. The effect for textured regions is hard to evaluate; in some cases it may be helpful, but it seems unlikely to work alone.

No justification is given for the values in the convolution mask. For the purpose of detecting a step edge in the presence of Gaussian noise, it is not the most sensitive.

Not enough experimental data is presented to give any feel for performance on real images.

Machuca and Gilbert 1981

Machuca, R. and A.L. Gilbert, "Finding edges in noisy scenes," Pattern Analysis and Machine Intelligence, PAMI-3, no. 1, January 1981, p. 303.

Abstract.

The authors present an argument based on Fourier analysis against the use of gradient operators to detect edges, especially in the presence of noise. They propose instead a "moment operator" and claim it well suited to detect step, ramp, and roof edges. The operator is tested on a variety of synthetic images and its performance compared to the Sobel operator [Duda 1973] using receiver operator characteristic curves. In all cases it is superior to the Sobel operator. Finally, they consider the theoretical advantages of mean versus median estimators in the presence of noise and find the noise removal properties of the mean estimator better in a low signal-to-noise environment.

The Fourier analysis argument against edge detectors is based on the fact that gradient operators emphasize high frequency components of both signal and noise. Nothing special in the way of an image model is discussed; the basic point seems to be that gradient operators basically consist of the second-order mixed partial derivative (the transition from the discrete to continuous domain is accomplished in a flurry of hand-waving) which has a parabolic power spectrum (i.e. is the square of the product of the x-frequency, y-frequency, and power spectrum of the original signal).

The moment operator is used to calculate the center of mass in a small rectangular neighborhood around each pixel. If at each pixel, the value of the pixel is replaced by a vector from the pixel to the center of mass around that pixel, at step and ramp edges the vector will point across the edge in the direction of higher intensity. Thus these edge can be detected by thresholding on the length of the vector, while the direction of the edge by the perpindicular to the vector.

Roof edges are a problem because at the edge the length of the vectors is zero. To detect edges here, one must examine the rotation of the vectors in a small circular neighborhood around a pixel; at a roof edge the vectors on one side of the edge are rotated 180 degrees from those on the other side.

Some computational techniques to enable real-time implementations of the algorithm are discussed next.

These methods were tested on images of disks whose edges were step, ramp, or roof edges, to which Gaussian noise of assorted standard deviation was added. In high noise conditions the roof edge was easily blurred, so median or mean filtering was used in these cases as a preprocessing step. ROC curves were used to evaluate the performance of the moment operator versus the Sobel operator, and in all cases the moment operator was superior. The preprocessing step of mean or median filtering seems to improve the edge detection results even for step and ramp edges (although the authors never say directly that it was used on these edges, they seem to imply it). This observation prompted them to investigate this step theoretically, with the result that mean filtering is superior to median filtering at low signal-to-noise ratios and the reverse at high signal-to-noise ratios.

Macrenhas and Prado 1978

Macrenhas, N.D.A. and L.O.C.Prado, "Edge detection in images: a hypothesis testing approach," in Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78), Kyoto, Japan, Nov.7-10, 1978.

Abstract

A simple statistical model for images is used along with statistical decision theory to develop a hypothesis testing approach to edge detection. A square neighborhood of four pixels is tested to determine the likelihood of all the possible horizontal, vertical, and 45-degree edges, plus the likelihood of no edge at all.

There is not much more to this paper than what is stated in the abstract. The image model is a two-dimensional autoregressive one originally proposed by [Habibi 1972]. The model is mentioned only briefly and not very well explained, but it seems to postulate some simple relationship among square neighborhoods of four pixels, with a little randomness thrown in. It is not clear why this should be a particularly useful model; perhaps Habibi's paper is more revealing.

In each four-neighborhood, the number of possible edges is six: four at 45-degrees (each of which has three pixels on one side of the edge and one on the other), one vertical, and one horizontal. The optimal solution to the hypothesis testing is stated but is difficult to evaluate numerically. Suboptimal solutions and approximations are proposed. Everything here seems relatively straightforward and unexciting.

Experimental results include running the algorithm on a cartoon image with synthetic noise added and on a meteorological satellite image. The claim is that the method takes no more computational effort than gradient methods and that the results are better under noisy conditions.

Marr and Hildreth 1979

Marr,D. and E.Hildreth, "Theory of Edge Detection," AI Memo 518, MIT AI Lab, April 1979. Also Proc.R.Soc.Lond.B., 1980, 207, 187-217.

The authors are concerned with finding a smoothing filter which will analyze the visual input into a number of channels related to physical scale.

They argue that such a filter should operate over a subrange of scales — not over all scales possible in the image. Furthermore, it should be spatially localized. From this they infer the (contradictory) requirements that the filter be both band-limited and space-limited. Although space limiting clearly follows from the localization requirement (and any realizable spatial domain filter is of necessity space-limited), the band-limiting conclusion is on shaky ground, since the idea of "scale" does not inexorably lead to the frequency domain. No arguments are presented to bolster the desire to consider the frequency domain. The reason that frequency domain methods work in engineering is the fact that exponentials are eigenfunctions of linear translation invariant operators, so one can use superposition to combine the effects of various bandpasses. Related is the convenient fact that convolutions are mapped to multiplications. The work under consideration uses exclusively linear methods, but does not present such an argument. On the other hand, if one uses nonlinear methods, there is no such justification. Indeed, this very problem occupied a large part of Norbert Wiener's later career and led him to his investigation of Brownian motion and measures on infinite dimensional spaces: he was interested in an analog of Fourier analysis for general nonlinear systems that would permit a similarly facile calculus to that which the latter provides for linear systems.

They argue further that the conflicting requirements of space- and band-limiting are optimally reconciled by minimizing the space-bandwidth product. For the appropriate definition of these terms, it is well known that the Gaussian (e^{-kx^2}, for the right k) achieves the minimum, so the authors conclude that the filters they want are Gaussian. Unfortunately, even if one accepts the doctrine of band-limiting, it is by no means clear that the Gaussian is optimal. In the first place, the Gaussian is neither band-limited nor space-limited. When one, say, bandlimits by truncation, it is no longer optimal. If one requires a strictly band-limited or space-limited function, i.e. one which is 0 outside of a given interval in either the spatial or frequency domain, the work of Landau, Pollak, Slepian [Slepian 1961, Landau 1961, Landau 1962] (pursued for this purpose, incidentally, by [Shanmugam 1979]) suggests that the optimal function is a prolate spheroidal wave function. Furthermore, these functions are eigenfunctions of the finite Fourier transform, which is significant since data are presented on a finite grid, a matter which the current authors don't address (the Gaussian (with the right k) is an eigenfunction of the continous Fourier transform). It may well be that Gaussians are good approximations to the optimal functions under certain circumstances. It is quite likely that literature exists concerning itself with that question; it is a subject worth pursuing for those interested in this approach.

The authors are interested in finding points of maximum directional derivative as edge locations, and they choose to locate these as zero crossings of a 2nd derivative. Based on cost considerations they opt for an isotropic 2nd derivative operator, the Laplacian ∇^2 (the only such), and wish to compute $\nabla^2(G * f)$, where G is the Gaussian. Since $\nabla^2(G * f) = (\nabla^2 G) * f$, they want to convolve with $\nabla^2 G$, which they approximate as a difference of Gaussians (DOG).

Logan's theorem (reconstructibility of 1 octave bandpass signals from their zero crossings) is invoked to justify use of zero crossings. However, the theorem is applicable only for 1 dimension, and the signals involved here have a bandpass of nearly 2 octaves. An argument is made that slope information may be adequate to bridge the gap (in analogy to the situation for the sampling theorem). On the other hand, there is no reason why reconstructibility should be a criterion, since there is never any requirement that an image *understanding* system should be able to reconstruct the input signal.

It appears that the authors are concerned that the zero crossing direction be perpendicular to the direction of "maximum slope of the directional derivative." Apparently, what this is supposed to mean is that $\nabla D_v^2 f$ should be collinear with v (where D_v^2 is the 2nd directional derivative in the v direction, the second derivative of a section of f taken along a line in the v direction, which can be written as $v^T H v$, where H is the Hessian matrix $[\partial^2 f/\partial x_i \partial x_j]$. Note that $D_v^2 f = D_v(D_v f)$, where D_v is the directional derivative in the v direction). Unfortunately, since they have a special case in mind where $\partial f/\partial y = 0$, they develop some strange ideas about the conditions and ramifications of their maximum slope concern. They state and claim to prove a theorem to the effect that the condition (confusingly stated) obtains if and only if $\partial f/\partial y = constant$. They fail to consider the actual slope in the if part, although it is true (albeit under the additional unstated condition that the slope is nonzero, i.e. that the 2nd directional derivative does not have a critical point at its zero crossing, e.g. if $f(x,y) = x^5$, $D_x^2 f(x,y) = 20x^3$). The only if part and its "proof," however, are totally fallacious; the purported proof shows only that $D_v^2 f$ may not be 0 for $v \neq \hat{x}$. Furthermore, it appears from the confusion around orientations of maximum slopes and zero crossings that the authors may not be aware that ∇g is always perpendicular to the locus of zero crossings of g.

The authors assume that coincident zero crossings from a set of contiguous channels imply a real edge and conversely. This so-called ***spatial coincidence assumption*** is not well-supported by any argument. The only situation for which it really makes sense is that of a very sharp edge between fairly large constant areas. Otherwise, it seems perfectly reasonable to believe that the edge will be visible at only 1 scale, while smaller scales will have inadequate sensitivity, i.e. their zero crossings will be essentially random, and larger scales may include other features, so their zero crossings will depend (arbitrarily) on those features as well.

Evaluation

There are no convincing arguments that the $\nabla^2 G$ or DOG operator is optimal or otherwise privileged in this context. As a 2nd derivative operator, though, it *is* esthetically pleasing because of smoothness and the Fourier domain symmetry (i.e. the Gaussian is an eigenfunction, or Gaussians in general are an invariant subspace of the Fourier transform). Zero crossings are a useful way to locate edges, but none of the "mathematical" or heuristic arguments presented about them here are convincing.

One must regard the assumptions and techniques of this work as tentative and experimental, rather than as a well founded theoretical or practical system. The ideas are based on intuition, perhaps good intuition, but lacking better justification must be regarded as only intuitive. The professed purpose is an explication of human vision (it is not clear how this is different for these purposes than say, vertebrate vision). Unfortunately, so little is known about human vision (e.g. there is no viable theory of how any but the most rudimentary information is coded or utilized), that one cannot draw any conclusions about the validity of any theory purporting to explain human vision, and in any case it is not our purpose to do so here. It is fair to say that the theory presented here is not obviously ruled out, but neither is it clearly the best or only possibility. As far as its being a theory of visual information or and engineering design, one can only say that it is an interesting provocative hypothesis, but not inexorable or proven.

Martelli 1972

Martelli,Alberto; "Edge Detection using Heuristic Search Methods," Dept of EE and Computer Science, NYU, University Heights, Bronx, NY, 10453. Also Computer Graphics and Image Processing, 1, 1972, 169-182.

Martelli,Alberto; "An Application of Heuristic Search Methods to Edge and Contour Detection," Instituto di Elaborazione della Informazione del Consiglio Nazionale delle Richerche, Pisa, 1973. Also in Comm. ACM 19, 1976, 73-83.

Following Montanari, he suggests that heuristics should be embedded in a figure of merit (FOM) rather than in code. But it is questionable that an FOM is enough in the way of heuristics — especially if it is not based on an analysis of real images.

Shows that any dynamic programming problem can be posed as a minimal path in a graph problem. Argues that this is good because the use of heuristics to speed up search in a graph is well-studied. However, the equivalence result is basically trivial (each variable expands to a set of nodes, one for each value). The advantage of dynamic programming is that it is far cheaper than graph search, and a better question is usually whether a graph search problem can be cast as a dynamic programming one. One can apply heuristics in the dynamic programming paradigm as well.

The variables x_i of the dynamic programming problem $FOM = f(x_1, \ldots, x_i, \ldots, x_n)$ are the edge elements — discrete valued and thus hard to generalize to continous θ edgels.

The figure of merit is defined in the form $FOM = \sum c_i(x_i, \ldots, x_{i+k})$ — see the criticism of Montanari that monotonically related c_i's don't lead to the same optimum. In this connection, no discussion of robustness with respect to FOM's is presented.

He derives a search graph for the dynamic programming problem, then uses algorithm A* to search the graph. The search graph is just a directed graph where the nodes are possible edge elements (pixel adjacencies) and the directed arcs are the allowed edgel successors.

The computation cost varies with S/N, since that determines the amount of searching which must be done. This is presented as a positive feature.

He uses a pairwise FOM of edge strength (nearest neighbor difference), but suggests that a larger local operator would improve noise performance.

Evaluation

Summary evaluation:
The technique is susceptibile to the standard problems associated with FOM.
The local operator is still very important.
The same problems as in [Montanari 1970], [Montanari 1971] are present.
The results look reasonable, but no results are presented for real images.
Analysis is required to decide whether the process can be made parallel — as it stands it is intrinsically sequential.

The technique presented by Martelli in this paper is not usable in its current form. With an appropriate local operator, reasonable FOM, the right discrete variables (i.e. edgel parametrization), it might produce reasonable results. But all that says is that global edge finding can be approached as a search problem. Furthermore, it seems likely that parallel search methods would be cheaper (as well as faster) than sequential, in analogy to simultaneous backward and forward searching in classical search problems. An intriguing idea is to use geometric information (i.e. relative direction) of other growing edges to compute the heuristic function (i.e. expansion ordering) for the search problem: edgels would be tried first that led toward something they might mate with.

Montanari 1970/71

Montanari, U.; "On the Optimal Detection of Curves in Noisy Pictures," Artificial Intelligence Laboratory, Stanford University, Memo AIM-115, 1970.

Montanari, U.; "On the Optimal Detection of Curves in Noisy Pictures," Comm. ACM 14, May 1971, 335-345.

The author presents a nonserial dynamic programming approach to find optimal 8-connected paths of a fixed length on a grid, and suggests a generalization which permits arbitrary length curves. Examples are displayed with mean square noise = mean square signal ($S/N = 1$) of length 45, with good results, though the examples are not related to real images. "Optimal" is with respect to a figure of merit (FOM); he uses one based on $\sum$ ***intensity*** $-\sum$ ***curvature*** (he is primarily interested in curves in the character recognition domain).

Evaluation

Use of this method as an edge detector requires developing an appropriate FOM. This is difficult, unless there is a canonical FOM imposed by the problem, since an FOM is not robust in the following sense. Viz., FOM's which are monotonic functions of each other (and as regular as you like) can give different global optima. In this case, to the extent that one can estimate the probability that local data were caused by an edge, one can use an FOM based on the entropy or information of the curve, so there is promise.

The requirement that the curve be an 8-connected path on a grid is troublesome, since one would prefer smooth curves as solutions. There is no easy way to translate the optimal path to a set of parameters representing a smooth curve, aside from an independent fitting process. Also, it is difficult to take into account any but the most local properties of the curve one is fitting, if for no other reason than the dimension of the interaction graph for the dynamic programming problem becomes prohibitively large.

Although one is guaranteed an optimum for the FOM, it is not certain that one necessarily wants such an optimum for image understanding applications, at least if the FOM is totally decomposable into spatially local components. The curve one is looking for is one which is the most meaningful in the context of the entire image intensity function (and world knowledge, set, etc.), and this meaning may depend on data away from the curve, which would lead to an intractable interaction graph for a naive implementation (i.e., one without special processing for these other parameters).

The dynamic programming approach is computationally very efficient; generalizations and adaptations of Montanari's method are probably worth pursuing, although it is not a trivial matter to do so.

Nevatia 1977

Nevatia,R.; "A Color Edge Detector and Its Use in Scene Segmentation," IEEE Trans. Systems, Man, and Cybernetics, vol SMC-7, no 11, November 1977, 820-826.

The goal is to define a Hueckel operator for the 3-color domain.

A review of color space representations is presented.

He claims there are 3 ways to look for color edges:

1) Choose a metric in the color space and look for discontinuities

2) Choose a basis and look for edges in the projection to each basis element separately

3) Do 2) but require uniformity to use 3 components together

He chooses to do 3).

However, what he actually proposes doing is minimizing the sum of the squares of the errors of the individual color component Hueckel fits. This is exactly equivalent to choosing an inner product on the color space such that the 3 color components are all orthogonal, then using the metric induced by the inner product, i.e. the Euclidean metric. This, as he points out, is equivalent to minimizing the individual components separately. Doing so, though, would lead to 3 fits for the 3 components which might have nothing whatever to do with each other (since one is not looking for the single edge that leads to all the data, but independent edges for 3 sets of data. Therefore, he imposes the additional constraint that the inclination angles for all 3 solutions must be the same, i.e. he adds the 2 equations $\alpha_1 = \alpha_2 = \alpha_3$. However, computing this angle is not easy, so instead he takes a weighted average of the 3 independent solutions (i.e., without the single angle constraint).

The idea of "best" fit implies a metric, since one must have a way to measure how good the fit is. Hence there is no way to avoid (explicitly or implicitly) choosing a metric for the color space. From this point of view, 3) is not really possible (unless one wants to be ad hoc), 2) is unsophisticated (though it may be adequate in many cases, but won't maximize S/N). It might be computationally efficient to choose not a basis, but a larger spanning set. 1) therefore is the way to go. Note that there may be more than one metric which is worth using simultaneously. It is also worth investigating the differences between using a metric such as

$$d(p,q) = ||p-q|| = \sqrt{\sum (p_i - q_i)^2}$$

and using a function

$$d(p,q) = ||p-q|| = \sqrt{\sum (p_i - q_i)^2}$$

$$\rho(p,q) = |||p|| - ||q||| = |\sqrt{\sum p_i^2} - \sqrt{\sum q_i^2}|.$$

Notice the latter is like the intensity difference.

In the connection of color, the following facts should be noted. A single color component is the intensity recorded through some filter, and corresponds to incident power. If the set of filters used are mutually orthogonal (and for the purposes of this discussion perfect transmitters in their passbands), then the total intensity or power is the sum over the individual components, i.e.

$$P = \sum P_i$$

If the P_i's are taken as the components, then this is not a Euclidean norm, which might be simpler computationally, but is more difficult mathematically, because it cannot come from an inner product. A possible approach to consider is to define $S_i = \sqrt{P_i}$ to be the coefficients of the basis vectors, so that one obtains

$$||S|| = \sqrt{\sum S_i{}^2} = \sqrt{\sum P_i}$$

which is a Euclidean metric. This can be thought of as using amplitudes rather than powers.

A second consideration regards the "true" orthogonality of the color components. Enough is known about the physical properties of light in modern times that one can define a spectral transmittance function R for a filter, or absorbence function for a detector. If the spectrum of the incoming radiation is given by $F : \mathbf{R}^+ \to \mathbf{R}^+$, then the response of the sensor will be given by $\int R \cdot F dx$. In this case, orthogonality of color components is the same as orthogonality of their spectral response functions in L^2. As is well-known, the passbands of the cones in the human retina have considerable overlap, and the same can be said of many filtering arrangements. Therefore, given a particular set of filters, the correct way to proceed is to use information about their spectral responses to obtain an orthonormal basis for the color space they define, as linear combinations of the amplitudes associated with the filters. Then the metric arises quite naturally from the physical properties of light.

Incidentally, this raises an interesting question. Suppose the filter characteristics are not known. Is it possible to derive an orthonormal basis from picture data alone? Presumably one would have to assume some kind of statistics for the actual spectral distributions of the subject matter of the picture data. This would shed some light on human perception of color, e.g. on the question whether knowledge of the form of the cone spectral responses (they are not similarly shaped, and some are quite wide) is hard-wired or can be "learned" (more exactly, compensated for).

O'Gorman 1978

O'Gorman,F., "Edge Detection using Walsh Functions," proc AISB p 195, July 1976. Also: Artificial Intelligence 10, 1978, 215-233.

O'Gorman shows that finding edge direction by fitting a plane and then taking its gradient direction is subject to systematic error for perfect step edges centered in a square window. However, this is a consequence of the shape of the window — a circular window would not have the same problem. Nevertheless, the analysis is salient because pictures are sampled on a square grid and rectangular operators are common.

He uses the 2-dimensional Walsh functions (tensor products of square waves) as an orthonormal basis for representing the image function. In analogy to [Hueckel 1971, Hueckel 1973] he does an L^2 (least squares) fit of a perfect edge on the first 6 terms (in his Walsh expansion).

The contribution of this idea derives from the fact that the Walsh basis bears a simple relationship to the digitization process (if one assumes square pixels). In particular, if one assumes that the digitization process takes place by averaging over a square pixel sized window, i.e.

$$g_{ij} = \int P_{ij} f dA,$$

where

f =image irradiance
P_{ij} =unit 2-dimensional pulse at the point (i, j)
g_{ij} =the sampled output,

then the P_{ij} constitute an orthonormal set whose span is identical to the Walsh functions of order less than IJ (where I, J are the cardinalities of the i, j sets). The higher order Walsh functions describe exactly only what goes on within pixels, which is precisely the information lost in the digitization process, so one has a perfect match of model to data. The Walsh basis differs from the single pixel basis most notably because the support is spread over the entire region of interest, i.e. the Walsh basis has *global support.* Truncating the series therefore results in global degradation, rather than local as would be the case with the analogous action of leaving off some set of pixels.

Unfortunately, incorporating sufficient Walsh terms to utilize all the picture data is equivalent to doing a fit of a perfect edge to the sampled data with the pixel value = average intensity assumption. This becomes extremely complex as the number of pixels increases, and if lateral displacements of edges are permitted, since the discontinuous pulse convolution kernel forces independent examination of numerous cases corresponding to the edge configurations' relations to corners of pixels. O'Gorman already has to consider 2 such cases for a 4x4 operator and 6 Walsh functions. As the space grows larger, so does the complexity, so that [Abdou] chooses to do an exhaustive search as his method of fit.

The advantage of everywhere differentiable functions (such as [Hueckel 1971, Hueckel 1973] uses) is that the lack of discontinuity permits a single set of equations to express the optimization problem. Of course, if one assumes a discontinuous sampling kernel, such expansions cannot avoid truncation error. However, if the sampling kernel is taken as (say) C^{∞}, its linear combinations (i.e. the functions $\sum \alpha_i \psi_i$, where the ψ_i are discrete translations of the sampling kernel ψ) become prime candidates for an expansion series. These can be frequency (or sequency) ordered. Expansion in terms of such functions has been extensively studied; use of truncated series fitting is worth investigating.

Ohlander 1975

Ohlander,R.B.; "Analysis of Natural Scenes," Dept of Computer Science, Carnegie- Mellon Univ, April 1975. (PhD thesis)

The author does region growing based on analyzing histograms of 9 color image parameters: the 3 raw R, G, B values, as well as the derived parameters of intensity, hue, saturation and the Y, I, Q parameters used in color signal coding techniques. In addition, values of their gradient as found by a Sobel operator are used, as is the local density of points above threshold in the gradient picture, called the "business matrix." He performs shrinking and expansion on the business matrix to eliminate thin regions (i.e., non-texture edges). The histogram analysis is based on a simple heuristic, and sometimes is done with manual intervention. Regions are found by thresholding.

Evaluation

Thresholding histograms ignores geometric relations in the data (a random permutation of the position of pixels doesn't change the histogram) as well as the photometric properties of the real world. Even so, use of 9 1-dimensional histograms is somewhat naive, since the pixel space is only 3-dimensional. It would be more systematic to use clustering techniques (which have an extensive literature) instead of 9 arbitrary 1 dimensional projections. (it would be more meaningful to think of color as a vector in $\mathbf{R}^3$)

This method can be expected to work on images that happen to be amenable to it, i.e. ones where the regions are pretty much homogeneous and separable from others by histogramming. In a more erudite approach of clustering, this amounts to being separable by one of 9 planes in $\mathbf{R}^3$, not even by any plane; and the latter is already known to be an overly restrictive condition for most clustering problems.

Pavlidis 1977

Pavlidis, T., Structural Pattern Recognition, Springer-Verlag, 1977.

Abstract.

Only chapter 2, entitled "Mathematical Techniques for Curve Fitting," and part of chapter 7, entitled "Analytical Description of Region Boundaries and Curves," are reviewed here.

Chapter 2

The problem, given an ordered set of discrete points in the plane, is to construct a smooth function "fitting" the points as a linear combination of a set of basis functions. The basis functions can take any form as long as they are linearly independent.

The quality of the fit can be measure by the integral square error E^2, equal to the sum of the squares of the differences between each point and the corresponding point on the approximating curve, or by the maximum error E^∞, the maximum of all absolute values of such differences. Many approximation techniques minimize either E^2 or E^∞; those minimizing E' are called "uniform." Uniform approximations can be drastically affected by a single point. Moreover, closed form expressions exist for the E^2 norm but only iterative methods for E^∞ one.

The method for calculating the basis function coefficients using the E^2 norm is then derived. The use of orthonormal basis functions is shown to simplify greatly the calculation.

The Karhounen-Loeve transform is discussed as an example of basis functions adapted to a specific problem leading to an accurate expansion with fewer terms. The basis functions are derived from a statistical sample of the data to be fitted.

Next methods for calculating uniform approximations are discussed. It is pointed out that E^2 approximations suffer from the drawback that they may miss such details of the data as sharp narrow peaks. Uniform approximations do not have this problem but are more expensive computationally. The problem of finding a uniform approximation can be posed as a linear programming problem.

A fundamental property of uniform approximations called equioscillation is mentioned: the maximum error is achieved at at least m+1 points with alternating signs (where m is the number of basis functions).

If the basis functions have the Chebyshev property (linear independence for any choice of m+1 points, e.g. polynomial and trigonometric functions), the revised simplex method of solving linear programming problems may be used; a detailed algorithm implementing this method is presented.

Because it is less obvious in vision than in speech or in electronic circuits that the natural process generating the signal can be thought of as a superposition of other processes, space domain techniques may be more relevant to vision than transform techniques. The theory of splines is relevant here because intuitively a natural scene seems more usefully divided into relatively homogeneous regions with possible sharp discontinuities between them, and splines are designed to approximate just this sort of function.

The discussion of piecewise approximations and splines deals with data in the form of continuous functions of one variable and with polynomial basis functions only. A spline is defined in the following way: the interval on which the data is defined is partitioned, and the points of partition are called knots. From each knot to the one following there can be a different set of coefficients on the basis functions; at the knots, equality of the first n derivatives (starting from 0) of the approximating functions to either side of the knot is usually specified to constrain the approximation.

Returning to discrete data for a moment, he points out two uses of splines for their representation. Interpolating splines have the knots coinciding with the data points and use as many continuity constraints as possible. They are useful for display purposes where a smooth curve can be shown instead of discrete data points. When the data are noisy, an interpolating spline is usually very oscillatory, and an approximating spline may be useful. But here a partition must be chosen; it turns out that often the choice of the knots is crucial to the usefulness of the spline, but leaving them as free parameters in the spline-computing process is "a very nasty mathematical and computational problem."

Calculating approximating splines with fixed knots is straightforward because between the knots the fitting problem is the same as that described above, except for the continuity constraints at the knots. He points that surface approximation by splines is a much less tractable problem than surface approximation by polynomials.

Piecewise approximations with variable knots are discussed next. Simple techniques are not useful. Consider random noise superimposed on a series of steps. Differentiation cannot be used to locate the breakpoints because of the high frequency noise. Low-pass filtering removes the noise at the expense of smoothing the steps. The problem of locating the knots is nonlinear, and the only available methods are iterative. Traditional descent methods do not work very well in this context because a local rather than global optimum may be reached, and because it may be necessary to estimate too many splines. He considers mainly piecewise linear approximations because they seem particularly relevant to shape perception and are analytically tractable.

The balanced error property is described and shown to be optimal for uniform approximations.

Some necessary conditions on error distributions for the optimality of L^2 approximations are stated and proved.

Several algorithms for piecewise approximation with variable joints are given: one for optimal joint location for uniform approximation, one for optimal join location for L^2 approximation, and one for optimal joint location for uniform approximation by descent.

Chapter 7

The first few sections are on projections and Fourier analysis and description of region boundaries and will not be reviewed here.

Next is a survey of boundary encoding techniques: chain codes, fitting of constant curvature arcs, fitting sequence of arbitrary curves à la chapter 2. Again the segmentation of the boundary into parts to be approximated by a single curve is the hard part. He divides approaches to this problem into four categories: search the boundary for the longest segment where approximation by a curve gives a sufficiently small error (similar to region growing); subdivide large arcs into small ones until a sufficiently small error is achieved; plot tangent versus arc length functions to identify sharp corners; use generalized Hough tranform methods, which are not discussed further because of their excessive computational cost.

The importance of curvature and corners is stressed, along with the effect of noise in making corners and points of high curvature more difficult to detect. He advocates piecewise linear approximations with variable knots as a technique for locating "true" curvature maxima (as opposed to those created by noise) and gives some theoretical reasons for doing so. A simple algorithm implementing this technique is presented.

Following is a discussion of merging schemes for approximating the boundary of a polygon. First is an algorithm to find a minimal length boundary with error at any point less than some threshold; it falls into category one above, and the algorithm is presented in detail. Assorted others. Next, some algorithms for the same purpose but in category two (splitting) are described. Finally, an algorithm combining splitting, merging, and breakpoint adjustment is described.

Prewitt 1970

Prewitt,J.M.S.; "Object Enhancement and Extraction," in Picture Processing and Psychopictorics, B.S.Lipkin and A.Rosenfeld,Eds., Academic Press, New York, 1970.

The paper is concerned with the entire image understanding problem:

image formation
image restoration
enhancement (including edge enhancement)
object extraction

The author provides a fairly extensive bibliography (237 references) of literature at that time (much of which is still germane).

The work is fairly sophisticated mathematically. E.g., Prewitt considers the Laplace, Mellin, Fourier, and Hankel transforms, moments, Haar-Walsh functions (cf. [O'Gorman 1976]), Chebyshev polynomials, point spread function (PSF), line spread function (LSF), edge spread function (ESF), modulation transfer function (MTF), and phase transfer function (PTF). She also discusses resolving power and restoration, including "super-resolution" for convolution degraded images (referencing e.g. Slepian, Pollak, Landau and applications).

Edge enhancement

A section devoted to edge enhancement discusses the gradient, generalized derivative, Laplacian, and discrete approximations to gradient.

As one means of obtaining an estimate of the gradient, she introduces the 3x3 (now so-called) "Prewitt operator:"

$$\begin{matrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{matrix} \qquad\qquad \begin{matrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{matrix}$$

This is used in a method of estimating the gradient by fitting a quadratic surface to data on a 3x3 square. The masks give $\partial/\partial x, \partial/\partial y$ for that surface directly from the data. This is exactly the method used by [Haralick 1980] for facets. Similarly, one can use the same idea for a 4x4 fit or a Laplacian.

She also discusses oriented edge masks, e.g.

$$\begin{matrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ -1 & -1 & -1 \end{matrix}$$

as approximations to the gradient ("compass gradient"), and gives some examples of their use.

A discussion of modified "crispening" (Laplacian) operators is presented, as well as of line enhancers (which are basically templates, i.e. matched filters).

Frequency filtering

Low, high, and band pass filtering is considered.

She discusses templates, matched filtering, and cross correlation for feature detection.

A good discussion of thresholding is presented.

The paper is an excellent overall survey of the then-existing methods for feature extraction, and in particular edge detection. By and large, the intervening 10 years have seen only minor improvements, so the analysis she presents is still relevant today.

Evaluation

Aside from frequency domain filtering, the methods presented, including the "Prewitt" operator, are completely local with small support — in her words, "context-insensitive." Consequently, global structures cannot contribute to the edge finding process, and the derived image description is limited to 1 or 2 local parameters which provide inadequate description of the image intensity function for all but especially regular images.

Unlike most gradient estimation or template matching operators, the Prewitt operator is based on well-defined process—the best fit of a plane. The gradient by itself is not sufficient for edge detection, since no discrimination is made between smooth and abrupt transitions, although plane fits can be used in more sophisticated ways (see e.g. [Haralick 1980]).

Roberts 1963

Roberts,L.G.; "Machine Perception of Three-Dimensional Solids," in Optical and Electro-Optical Information Processing, J.T.Tippett et al., Eds., Massachusetts Institute of Technology Press, Cambridge, Mass., 1965, 159-197. Also Technical Report No. 315, Lincoln Laboratory, M.I.T. (May 1963).

The research described seeks to match a narrow class of prismatic solids to models, starting from raw picture data. There is a wide range of issues which the author had to address to achieve this; since we are concerned here with segmentation, we ignore most of the other contributions of the paper.

The central task the program performs is to match a wire frame model to derived wire frame data. An important part of this consists of vertex matching. To this end, he tries to fit n-point data (2 dimensional) to an n-point model (3 dimensional) by finding the best transforms H, D in homogeneous coordinates such that

$$AH = DB + \epsilon,$$

where

A =n points (x, y, z, w) from the model
B =n points (y, z, w) from the data (uses x as projection axis)
H = 3x4 homogeneous perspective transform
D = Diagonal nxn scale matrix
ϵ = error matrix

He solves this as a least squares problem.

Local edge detector

He first takes the square roots of the pixel values, on the basis of psychophysical evidence which he cites. In a 2x2 pixel window, let the square root values be

$$\begin{matrix} a & b \\ c & d \end{matrix}$$

The edge measure is then defined by

$$\varphi = \sqrt{(a-d)^2 + (b-c)^2}$$

This is proportional to the gradient magnitude of a least squares fit plane (e.g. [Haralick 1980]). I.e., if F is the best fit plane,

$$|\nabla F| = \frac{1}{\sqrt{2}}\sqrt{(a-d)^2 + (b-c)^2}$$

Roberts cautions that his line finder "makes mistakes in complex pictures and is a complex special-purpose program demonstrating very few general concepts." One must keep in mind that this was a pioneering work and his main interest was higher level model matching.

We summarize the operations performed in line finding in the following lists.

Edge detection process

$\Phi = R_\nabla(P)$ (do Roberts cross operation, i.e. compute $|grad|$)
take max on each 4x4 square of a tesselation
threshold
correlate (i.e. sum) along lines of length = 5, for values of $\theta = n \cdot 45°$
threshold on the ratio $\frac{best}{worst}$ of the line values, yielding edges

Linking

- connect edgels if:

    1) they lie in contiguous 4x4 squares

    2) they are related by a < 23° change in direction
- eliminate singletons
- apply an ad hoc cleaning processes for small triangles, quadrilaterals, and spurs

Curve representation and segmentation

- least squares fit straight lines to linked sets.
- uses sequential (updating) method of fit
- first done on connected edgels
- choose a random starting point, then proceed until:

    1) a branch is reached, or

    2) an error threshold is exceeded for the line fit, in which case back up until the local angle to the fit is cut by 1/2.

The remainder of the paper is concerned with the recognition and display of polygonal 3 dimensional objects.

Evaluation of line finding

This was an early effort. It probably is not bad for straight lines, though it seems to miss a lot. Curved edges or complex scenes are not handled. Many ad hoc methods are used.

The technique presented here has no hope of working where there are wide variations in smooth shading gradients, since the thresholds are global, and the gradient operator cannot discern whether the signal is from a smooth gradient or a local step.

Of course, it must be stressed that Roberts broke ground in the use of his gradient operator, as well as in the use of homogeneous coordinates, the fitting of 2 dimensional data to 3 dimensional models, and in line following.

Rosenfeld, Hummel and Zucker 1976/77

Rosenfeld,A., R.A.Hummel, S.W.Zucker, "Scene Labelling by Relaxation Operations," Computer Science Center, Univ of Md, TR-379, May 1975. Also IEEE Trans. Syst. Man Cybern., SMC-6, no.6, June 1976, 420-433.

Zucker,S.W., R.A.Hummel, and A.Rosenfeld, "An application of relaxation labelling to line and curve enhancement," IEEE Trans. Computers C-26, 1977, 394-403.

The authors generalize a method first developed by Waltz for propagating constraints in a graph. Waltz called it "filtering" and used a sequential process; the present authors call it "relaxation" (not related to a method used for solving partial differential equations) and do it in an essentially parallel way. One starts with some finite set of objects, some set of interpretations for each object, and a graph where the nodes are the objects and arcs represent mutual constraints between interpretations ("labellings") of the objects. The authors treat 3 types of labelling sets: discrete (finite set of labels) fuzzy (finite set of labels with weights between 0 and '), and probabilistic (finite set of labels with weights between 0 and 1). A generalization to a continuous set of labels is not hard to imagine, and would be useful in the applications, for example to represent the orientation of an edge. For the probabilistic case, they readily show that the relaxation process has a fixed point, a necessary condition for convergence. They go on to show that for a class of linear operators with eigenvalues of norm no more that unity, convergence to the unique fixed point is guaranteed. Unfortunately, it's not an interesting case, because the fixed point is independent of initial conditions, i.e. input data. They also present a more interesting nonlinear operator, but are unable to prove that it converges. One can probably invoke one of many variations of the contraction mapping theorem to show convergence for their linear case as well as nonlinear mappings which are contractions in the appropriate sense, thereby expending less effort and achieving greater generality. The important point, however, is that a wide, useful class of such relaxation operators converges. One can even say something about the speed of convergence, based, for example, on the eigenvalues of the relaxation iteration operator. The idea is closely related to dynamical systems, which has interesting implications for neurophysiology and hardware design. If one views the state space as a free vector space on the labels over the field of weights (which we take to be **R**), then the relaxation is a map of that space to itself. If that map is a diffeomorphism, it may be embeddable as a time-one map of a flow, i.e. it may be the discrete time snapshot of a continuous dynamical system. In that case, the process can be performed by a system of ordinary differential equations. Incidentally, economies can be gained by transforming the state space so as to diagonalize, upper triangularize, or Jordan normal form-alize the relaxation map if it is linear or approximately so. Even if the relaxation is not embeddable in a continuous dynamical system, nearly all the machinery of dynamical system theory is available. For example, if the fixed points are known, theorems are available telling us under what conditions there is convergence near the fixed point, whether the system is stable (i.e. robust with respect to the choice of relaxation parameters), etc.

Unfortunately, the experimental results presented are not very impressive. However, it is plausible that that is because the continuous spectrum of labellings generalization has not been made, because the relaxation (compatibility) coefficients are chosen ad hoc without consideration of robustness (they use the word a few times, but don't have any notion that one can say quantitative things about it). The poor experimental results, therefore, should not be regarded as an indictment of the idea. Rather, it should be developed with greater sophistication. For example, one should consider the effects of noise in a quantitative way. One should try to discover whether there are any global quantities being optimized in the solution. One might consider generalizations to infinite sets of objects, e.g. curves. Thus the local label would be a probability density function for, say, edge orientations and strengths. This leads to an infinite dimensional state space, and although there is a respectable theory of dynamical systems in such spaces, one must confront the computational difficulties. However, since the function factors composing the space are on compact domains, there is a natural decomposition in terms of Fourier series (of the probability densities in the orientation-strength domain), and it is equally natural to truncate these series, so one again obtains a finite dimensional characterization of the state space. One would then want to study the relationship between such a process and, say, variational methods. One can expect that for reasonably regular (finite dimensional) systems, there will be a finite number of fixed points outside of a small neighborhood, hence only a finite number of final states. One should give some thought as to whether that is an acceptable situation. It could be remedied by altering the relaxation coefficients based on the current state.

Evaluation

In the form presented in these papers, relaxation doesn't seem to be a very good edge or curve detector, and its behavior is not too well understood.

Rutkowski and Rosenfeld 1978

Rutkowski, W.S. and A. Rosenfeld, "A Comparison of Corner-Detection Techniques for Chain-Coded Curves," University of Maryland Technical Report TR-623, Jan. 1978.

Abstract.

Several methods of detecting angles or "corners" on digital curves are defined. Results obtained using these methods are compared with each other and with subjective angle-detection judgments.

A digital curve is defined by a chain code, i.e. a sequence of linked vectors. Finding corners and estimating angles is complicated by the fact that those arising from the discreteness of the digitization must be separated from those representing more meaningful changes in direction of the curve. The authors present several ad hoc methods for detecting these meaningful changes.

The k-curvature method defines curvature at a point as the angle between two vectors, the first a weighted average of the k vectors before the point in question and the second that of the k after. This angle calculated along a curve will presumably peak at a corner, so non-maximum suppression can be used to detect corners sharply.

The arc-chord distance method draws a chord between the first and k + first point on the curve, calculates the distance from each point spanned by the chord to the chord, then draws a chord between the second and k + second point, etc. At point only the maximum distance between the point and a chord is saved. As before, the distance tends to peak at a corner, and non-maximum suppression can be used for precise localization. The authors mention that results with this method vary with the orientation of the corner, but their explanation why is not clear (to me, anyway).

The peak-finding method projects each point on a curve onto each of a set of unit vectors at different orientations. The result is a matrix consisting of the lengths of the projections, with the rows corresponding to points and columns to unit vectors. A peak-finding technique is applied to each column, and through some manipulations which are not clear to me, a curvature vector is obtained, effectively summing the results obtained from the various orientations.

The histogram overlap method creates two direction histograms for each point, one for the preceding k points and one for the succeeding k points. (The direction histograms plot the number of occurrence of each of a specified number of directions.) A measure of the disjointness of the two histograms can then be used as a measure of curvature. The sum of the absolute values of the differences between the two histograms at each direction is one such measure. Once again, nonmaxima suppression can be used to detect corners sharply.

The methods were tested by comparing the responses of humans asked to find the "turns" in a given curve and to rate them on a scale of 1-3, to the responses of the four methods described above. Methods 2 and 3 have a tendency to yield satellite peaks. Method 4 seemed less useful than the others.

Criticism and evaluation

All methods are totally ad hoc. No model of a curve or of curve corners is presented, so there is really no clear idea of the problem to be solved or why the problem is hard. It follows that analyzing the performance of the methods is difficult as well, although in the one case tested three of the four seemed adequate.

Shafer 1980

Shafer, Steven A.; "MOOSE. Users' Manual, Implementation Guide, Evaluation," Bericht 70, IfI-HH-B-70/80, Fachbereich Informatik, Universität Hamburg, April 1980.

Shafer describes a system following Ohlander's technique of image segmentation by the use of multi- spectral histograms. The implementation is essentially automatic, and reasonably fast (30 seconds on a PDP-10 to segment a 96x128 image, and 20-25 minutes total time with all displays). See the remarks about Ohlander's work regarding the histogramming technique.

The author himself provides some criticism of the technique. The main shortcoming pointed out is referred to as the "majority rule" problem, which occurs when the histogram peak separation process is dominated by large regions. In that case, if a small region happens to be situated in a narrow valley between the large regions (i.e. the large histograms nearly overlap), the small region will be broken in two arbitrarily. This is a consequence of the fact that histogramming ignores geometric relationships. The solution proposed is to first crop the picture so that a small region to be segmented from its surround becomes a large region in the sub-picture. Of course, this amounts to an approximate segmentation. No method is proposed to do this automatically, though the author argues that the cropping idea is robust by showing that including some other objects in the cropped area still allows reasonable performance. This seems to indicate that histogramming works better for very small pictures. A seductive idea (not suggested by the author) is to try arbitrarily subdividing the picture and simply segmenting the smaller pictures. Unfortunately, this will create non- trivial problems in merging regions across subpicture boundaries. In view of the many shortcomings of histogramming and thresholding techniques, it does not seem worthwhile to pursue improvements.

The author also points out the following problems. Many small areas at the boundary of a region are lost since the boundary is sensitive to the threshold. He suggests the solution of merging them after other segmentation is complete. Regions of non-constant intensity cannot be handled, i.e. the technique fails in the presence of any shading. Strangely, he points out that the gradient requires 2 parameters for description, but he does not know how to express this in "one-dimensional features." Presumably, he means he wants to histogram the gradient somehow, and to use Ohlander's methods means selecting a single parameter to histogram and threshold. In analogy, e.g. to Ohlander's use of $R+G+B$, gradient magnitude seems like a reasonable candidate for one such parameter, and it is unclear why the author neglects it.

The eventual goal of this system is for use in an object tracking system. One might hope that even if one couldn't overcome the problems of segmenting a single image, the segmentation would at least be stable from frame to frame. This seems to be a false hope. Thresholding can be thought of as creating boundaries where some level plane intersects the image parameter value function, so that different thresholds correspond to differen height contours on a topographic map. At boundaries with small gradient, geometry will change rapidly with threshold value; and at peaks, valleys, and saddles there will be a change in topology as a function of threshold value. If this function has lots of bumps, and if it is changing with time, then there is a serious problem of keeping track of what is going on. The problem becomes essentially equivalent to keeping track of the whole parameter function, since the "objects" created by the segmentation are quite likely not to be stable. Of course, ones which *are* stable *can* be tracked this way. What can reasonably expected to be stable? An object with regions of constant parameter value (shading is tolerable in a system which looks at hue, as long as hue is constant — an admittedly unlikely situation), moving through light in such a way that the reflectance changes very slowly relative to the motion, against a background having very different spectral characteristics, occurring in an image where everything else also has different spectral characteristics than the object and its background. This appears to be a very limited domain, though there may be useful applications, nevertheless.

Shanmugam, Dickey and Green 1979

Shanmugam, K.S., Dickey, F.M., Green, J.A., "An optimal frequency domain filter for edge detection in digital images," IEEE Trans Pattern Analysis and Machine Intelligence, PAMI-1, 39-47, January 1979.

The authors consider the 1-dimensional edge detection problem, with the proviso that "symmetries appropriate to the 2-dimensional problem are retained." Their goal is to obtain a frequency domain filter to concentrate maximal energy near an edge. The model for an input edge is the unit step.

More particularly, the authors require a strictly bandlimited filter (i.e. a filter whose Fourier transform has its support on an interval surrounding the origin), and they seek to maximize the power in some interval around the origin in the space domain for the filter output response to a unit step.

Following Landau, Pollak, Slepian [Slepian 1961, Landau 1961, Landau 1962] they decompose in terms of prolate spheroidal wave functions, and show that the optimal filter output is ψ_1, the order 1 prolate spheroidal wave function, with the space-bandwidth parameter dependent on the space and bandwidth cutoffs required. Note that this result is in contradiction to the assumption by [Marr 1979] that the Gaussian is "optimal" (i.e., that is proven not to be so under conditions of strict bandlimiting). Specifically, the transfer function of the optimal filter is given by

$$H(\omega) = \begin{cases} K \frac{\psi_1(\frac{\Omega I}{2}, \frac{\omega I}{2\Omega})}{iF(\omega)}, & |\omega| < \Omega \\ 0 \quad , & |\omega| \geq \Omega \end{cases}$$

where K is a constant, ψ_1 is the 1st order prolate spheroidal wave function, Ω is the bandpass interval width, I is the spatial interval width, and $F(\omega)$ is the Fourier transform of the ideal input. The only information used about the input and filter to derive this formula is the fact that they are odd and even functions, resp. There is no particular justification for requiring the filter to be even (it gives a neater result) except that a 2 dimensional generalization is rotational invariance.

Blurred edges are modelled as the difference of exponentials to obtain a symmetrical sigmoid function (only once continuously differentiable, though). They show that if the resolution interval I is larger than the blur width (defined by the 90% points), then the filter is still a good approximation to optimal in an appropriate sense.

A Gaussian noise analysis is also presented, showing that S/N improves with increasing space-bandwidth product, e.g. coarser resolution, not a very surprising result in view of many others to the same effect. An expression for S/N is given.

The experimental results are not very impressive when compared to nonlinear edge detectors (e.g. after thresholding), but they show a clear improvement over other standard linear filters, e.g. high pass, Laplacian.

Evaluation

This is not a direct method of detecting edges, but rather should be regarded either as an enhancement method, or, more importantly, as a possible precise approach that Marr and Hildreth could have taken in finding the optimum filter to reconcile space limiting and band limiting. However, unless one requires that computations be done in the frequency domain, there is no persuasive argument presented here or elsewhere that requires using a bandlimited filter as part of edge finding, so it is not at all clear that all this elegance is necessary or useful.

On the other hand, it would indeed be satisfying to learn that bandlimiting *is* required, so that the prolate spheroidal wave functions might be worth experimenting with, and should at least be kept in mind.

Shirai 1975

Shirai, Y., "Edge finding, segmentation of edges and recognition of complex objects," Proc. 4th IJCAI, 1975, 674-681.

Abstract.

An approach to recognizing real-world objects such as books or a telephone on a desk is described. The system consists of: edge finding from light intensity data, segmentation of edge into straight lines or elliptic curves, object recognition by matching lines to models. An example of locating a lamp, a book stand, and a telephone is shown.

The edge operator used is basically a gradient operator. Edge points are found and then linked in separate stage. I'll skip the details of the edge detection and linking and recognition phase because my primary interest is in the edge segmentation step.

The two steps in the segmentation phase are: (1) segment the edge into lines and curves, and (2) approximate each segment by a straight line or an elliptic curve. The method used in the first step is to define the curvature at every point on a curve by the angle between the line segment connecting the point in question and the point k points preceding it, and that connecting the point and the point k points succeeding it. Narrow peaks between regions of low curvature are knots between line segments, while broader regions of above-threshold curvature correspond to curved segments. Next undefined and curved segments are checked for possible mergers with adjacent segments. Finally, a straight line or curve is fitted to the set of edge points in each segment.

Deming's method is used to find the line or curve which minimizes the sum of the squares of the distance from the curve to each point (as opposed to conventional methods which minimize the sum of the squares of the distance only in the y-direction.

Criticism and evaluation

The approach is straightforward but ad hoc. There is no model of curves or of their corners, no consideration of where the errors in location of edge points arise (e.g. sensor noise, blurring, his edge point location algorithm).

Somerville and Mundy 1976

Somerville, C. and J.L.Mundy, "One Pass Contouring of Images Through Planar Approximation," Proc. of the 3rd International Joint Conference on Pattern Recognition (IJCPR-76), Nov. 1976 (IEEE 76CH1140-3C).

The authors state that they are interested in finding contours of the intensity function, but what they mean by this is finding places of large change of gradient.

Their first goal is to represent the picture data compactly for further processing. This is a good idea, since it is necessary to have a representation of the intensity surface for varying neighborhoods — not just single pixels. An important reason for doing so, which they do not mention, is to synthesize semi-global but accurate information. (By semi-global, we mean regions larger than a single pixel or pair of pixels, yet smaller, usually much smaller, than the entire image.)

The primitive regions to be used for region growing are triangles. These are initially formed by drawing diagonals for each set of 4 points so as to keep similar intensities together. The region growing is then done by a process of raster scan local merging. The merging criterion is as follows.

1) Compute the normal vector of the intensity function on the next triangle. This is not normalized — it is actually the 3 dimensional gradient.

2) Compare with the current *average* normal vector for each whole adjacent region.

3) Merge the triangle into the region if the magnitude of the vector error ($|\mathbf{n}_T - \mathbf{n}_R|$) is less than a threshold based on region size:

$$\epsilon_{max}(A) = k_1 e^{-k_2 A} + k_3$$

This can be criticized as follows.

2) Presumably, they do this because they want regions of uniform normal. But it seems more reasonable to compare normals *locally*, leading to locally uniform normal, i.e. regions of slowly changing normal.

3) The adaptive threshold is not well justified. The stated purpose is noise immunity — presumably, values for large regions should be more stable, so there are 2 terms, one for

the region noise, one for the triangle noise, though this is not explicitly stated. Since the gradient is a linear operator, one could in fact explicitly solve for the noise characteristics of the expected difference in normals. The region component would be of the form $\sigma = k\sigma_0/\sqrt{A}$, and in fact the standard deviation of error in normals is given by

$$\sqrt{\frac{k^2}{A} + \frac{c^2}{3}},$$

where k^2 is the mean square contribution of each pixel in the region to the expression for the normal, and c^2 is the analogous quantity for a single triangle. In this light, the threshold adopted by the authors is seen to be a linearization and exponential approximation to this function, for a fixed standard deviation of image noise. Furthermore, since the merging is don: on a raster scan, the merging predicate will result in different behavior near the tops of regions as compared to the bottoms. Not only that, but this can happen in a discontinous way, when 2 regions suddenly get merged.

The entire process is equivalent to edge detection based on computing a gradient from x and y first differences. However, the edge predicate is adaptive in the sense that the threshold is based on the mean gradient of adjacent regions (in this case, only of regions above, i.e. earlier in scanning). The adaptive part isn't a bad idea, but using an operator with a support of 3 will lead to noise problems, as well as problems with discerning larger scale features.

Experimental results

A single example on a 64x48x6 picture is given. A reconstruction of the original is presented, based on linear interpolation about the centroid of each region. This result is not impressive. The authors are concerned with data compression and reconstructibility, but from the point of view of image understanding, reconstructibility should, however, not be seen as a measure of performance. The region boundaries displayed, though, do not appear significantly better than other, local, methods. It would be interesting to see the results of a process incorporating the improvements suggested above, viz.

- gradients computed for larger neighborhoods
- thresholds based on picture noise and exact formulas
- merging based on local information. Alternately, one could iterate taking gradients.
- some isotropic merging process (which might result in the requirement for more than 1 pass).

Even so, the gradient idea leads to difficulties if an edge should pass through the operator support — one might get many regions perpendicular to the edge, which would be elongated along the edge, but broken up as the geometry of the edge changes — in other words, poor behavior. A plane is too simple a model for the local intensity surface.

Turner 1974

Turner,K., Computer Perception of curved objects using a television camera, Ph.D dissertation, Edinburgh University, November 1974.

We are concerned here only with the edge finding aspects of this work

The author gives a brief critical synopsis of earlier line finding work:

Binford-Horn
Griffith
Herskovits and Binford
Hough
Hueckel
Kelly
Murphy
O'Gorman and Clowes
Pingle
Pingle and Tenenbaum
Roberts
Shirai
Tenenbaum

The edge finder is very simple, Using first difference of adjacent pixels, followed by thinning, and further by a local tracker (inchworm).

A short review of curve segmentation is provided.

He discusses the (φ, s) representation of plane curves, defined by

$$\varphi = \text{tangent direction, and}$$
$$s = \text{arc length.}$$

Then

$$d\varphi/ds = \kappa \text{is the curvature, and}$$
$$d\varphi/ds = constant \Leftrightarrow \text{the curve is a straight line}$$
$$\Leftrightarrow \varphi = \text{a linear function of } s.$$

He finds curves by fitting straight line segments to the (φ, s) data.

Yakimovsky 1976

Yakimovsky, Y., "Boundary and Object Detection in Real World Images," Journal of the ACM, vol.23, no.4, October 1976, p. 599.

Abstract.

A computer solution to the problem of automatic location of objects in digital pictures is presented. A self-scaling local edge detector that can be applied in parallel on a picture is described. Clustering algorithms and sequential boundary following algorithms process the edge data to local images of objects and generate a data structure that represents the imaged objects.

He proposes a local edge detector for step edges (constant on either side of edge) that compares disjoint adjacent neighborhoods of pixels and uses hypothesis testing approach (the Neyman-Pearson criterion) to choose between two possible hypotheses: both neighborhoods come from the same Gaussian distribution, or they come from distinct ones. The maximum likelihood ratio is used as the estimate of edge strength.

To deal with roof-type edges, he assumes both sides of edge lie on distinct planes if there is an edge and on the same plane if not. In both cases superimposed Gaussian noise is assumed to account for any remaining randomness. The maximum likelihood estimates of the parameters for the plane and the gaussian distribution is computed for each neighborhood separately, then for both neighborhoods together. Again, the Neyman-Pearson principle is used to decide whether two planes are present or one. He has not yet implemented this method.

In both of these cases, edge location is determined no more accurately than to lie betweens pixels adjacent in the four-connected sense, i.e. there is no subpixel precision attempted.

Neighborhood selection for the hypothesis testing is admittedly totally ad hoc. He catalogues the different shapes he uses for regular edge, line, t-corner, etc, all of which were derived experimentally. The guideline for the size of the neighborhood is a "reasonable balance between noise and size of object."

Measures of edge strength similar to those above are proposed to evaluate confidence in parameters found by the Hueckel edge detector: "model driven confidence evaluation in the existence of the suggested edge."

Determining exactly between which two pixels the edge lies is nontrivial because the output of the edge detector along the normal to the edge often rises gradually to a poorly defined peak as the actual edge is approached. Local maximum selection is one remedy, but he points out that for practical reasons it is often better to treat the area around the edge as ambiguous. His implementation used neither option but rather left exact localization of the edge to the region grower.

He discusses in great detail the region growing methods that he has implemented, claiming that he introduces a new algorithm for clustering points into regions based on a search for "valleys" of edge values. As far as I can tell, all he has implemented and discusses is a single-pass scan-line algorithm which examines sequentially two-by-two neighborhoods with already computed edge data and uses local edge information, and region and boundary (linked edges) computed for the three upper/ left pixels in the neighborhood, to form, extend, and merge regions and boundaries in an obvious way. His algorithm is specified in an ALGOL-like language. He follows with a brief discussion of relatively trivial methods for postprocessing the output of his algorithm to improve the region segmentation, one of which uses statistical methods similar to those with which edges are detected. All these methods are based on local properties of regions rather than, for example, shape. A somewhat more global statistical method is proposed for merging regions.

A brief summary of what I see to be the weaknesses in his approach: the loss of most of the spatial information, particularly in the fact that that pixel values near the proposed edge are weighted no more heavily that those in the neighborhood further away from the edge; the ad hoc methods of neighborhood selection, which is actually another example of his failure to use spatial information; the lack of an explicit method for dealing with blurring, which is one aspect of his unrealistic model for edges, another of which is the assumption of constancy or planarity on either side of an edge; the failure to perform subpixel interpolation and the resulting loss of relevant information; a far-too-trivially-simple region grower, which does not use the edge strength along a boundary or other available information in its computations.

3.1.8 – Glossary

Adjoint (of an operator). The *adjoint* of an operator $A : L^2 \to L^2$ is the unique operator A^* such that $(Af, g) = (f, A^*g)$ for all f, g.

Diffeomorphism. A *diffeomorphism* is a differentiable homeomorphism with differentiable inverse. The degree of differentiability is usually implicit from the context.

Flow. A *flow* on a manifold M is a map $\psi : M \times \mathbf{R} \to M$, (with $\psi(x, t)$ usually written as $\psi_t(x)$) such that
(1) $\psi_0 = identity$
(2) $\psi_{t+s}(x) = \psi_s \circ \psi_t(x)$,
for all $x \in M$ and $t, s \in \mathbf{R}$. The terminology derives from the picture of flow lines, e.g. of a moving compressible fluid. A single flow line describes the motion of a single point of M as a function of the time t. Sometimes a flow is defined to have some smoothness properties, e.g. be a diffeomorphism, and in practice, it usually is.

Hilbert space. A Hilbert space is a complete normed inner product space.

Homeomorphism. A *homeomorphism* is a continuous 1-1 onto map between topological spaces whose inverse is also continous.

$L^2(\mathbf{R}^2)$. The space of all Lebesgue square integrable complex-valued functions on the plane, equivalenced by the functions of integral 0, with the inner product $(f, g) = \int fg^* d\mu$, where $*$ denotes complex conjugate.

Laplacian. The *Laplacian* is the 2nd order differential operator, also written ∇^2, denoted by $\partial^2/\partial x^2 + \partial^2/\partial y^2$, so that $\nabla^2 f = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2$.

Normal (operator). A is a *normal* operator if $AA^* = A^*A$, where A^* denotes the adjoint of A. Translations and rotations (in the plane) are readily seen to be normal.

Support. The *support* of a function $f : X \to \mathbf{R}$ is the closure of the set of points $x \in X$ where $f(x) \neq 0$. Put another way, it is the smallest closed set containing all the places where f does not vanish.

Time-one map. The *time-one* map of a flow ψ_t is simply the map defined by the flow at time 1, $\psi_1 : M \to M$.

Unitarily equivalent. Two operators A, B are said to be ***unitarily equivalent*** if there is a unitary operator U such that $U^{-1}AU = B$. Equivalently, A and B are related by an isometric change of coordinates.

Unitary (operator). An operator U on an inner product space is ***unitary*** if $U^* = U^{-1}$, where U^* is the adjoint of U. It is not hard to see that the unitary operators are the isometries.

3.2 TEXTURE ANALYSIS

(This section adapted by Ramakant Nevatia from his forthcoming book "Machine Perception," to be published by Prentice-Hall 1982, Copyright Prentice-Hall Inc.)

Surfaces of natural objects are not always homogeneous in a local attribute, e.g. intensity or color, but may have a more or less uniform observed pattern, called the visual texture, possibly generated by the physical texture, as in a rough wall surface, or simply the markings on a surface, as in a wallpaper. In some cases, it is natural to view a collection of objects as a single entity, e.g., a grass field or a wall of bricks. In these cases, the pattern of individual objects determines the texture of the collection.

Two common approaches to texture analysis have evolved. One considers texture as an instance of a random process and measures the parameters of this process. The other tries to characterize the primitives of the texture and their pattern explicitly.

3.2.1 - Statistical Texture Measures

The statistical measures have been motivated by a lack of simply described patterns in natural textures. Further, these measures have been supported by an important conjecture, due to Julesz [Julesz 1962], that humans are unable to distinguish textures that have the same second and lower order statistics, but differ in one or more higher order statistics. (n-th order statistics are determined by the joint probability distributions of n pixels at a time.) Julesz conjecture is supported by a large number of examples; however, several counterexamples have been found [Caelli 1978b],[Julesz 1978], [Purks 1977], [Victor 1978], [Gagalowicz 1980]. These counterexamples have patterns that have some discernible micro-patterns and the local second order statistcs are different from the global statistics.

Statistical measures are based on the average properties assumed to be invariant over an entire region. The number of suggested statistical measures is large and only the commonly cited and used approaches will be described here.

a) First Order Measures

The simplest measures are based on first-order statistics, i.e., probability distributions of single pixel attributes. Some examples are mean and variance of intensity. More sophisticated first order measures are based on histograms of the individual pixel attributes. These measures are not strictly texture measures, as they are not even dependent on the spatial distribution of the pixel attributes, but are still useful for many naturally occurring textures.

b) "Texture Energy" Measures

An improvement over the first order measures using pixel attributes is to detect the presence of certain features in the texture and then to compute the first order statistics of these features. The density of edges, detected by a local edge detector, is commonly used to distinguish between "coarse" and "fine" textures.

[Laws 1980] has generalized this concept to determining a variety of features by convolving the image with a variety of filter templates, and then measuring the "energy" of the outputs The filters consist of 3x3 or 5x5 masks and detect the presence of edges and lines in various orientations, and corner-like features. The filters were determined empirically by testing with some natural textures. The energy measures compute properties over a larger window (15x15). Simple measures are average and variance of the filtered outputs. As the various measures are not independent, the resulting features are combined into a smaller set.

c) Fourier Measures

As textures are viewed to be at least semi-periodic, the Fourier transform of an image window can be expected to have distinct peaks useful for texture discrimination. Bajcsy used filters in the Fourier domain, consisting of annular rings and strips in different orientations [Bajcsy 1973]. The outputs of these filters was used to generate symbolic descriptions such as blob-like, homogeneous, random, monodirectional and bidirectional. This technique was applied successfully to natural scenes containing textures such as grass, water, sand, trees, etc.

A disadvantage of the Fourier approach is that, except for perfectly periodic textures, the energy in the frequency domain is scattered, and similar peaks may be caused by a nearly periodic texture and a single strong edge.

d) Second Order Measures

Haralick et al. suggested a scheme for estimating second order joint probability densities and devised measures based on them [Haralick 1973]. The second order statistics are given by $P(i, j, d, \theta))$, the probability of a pair of pixels separated by a distance $|d|$ in direction θ having the intensity values of i and j. These statistics can be computed and stored in the form of co-occurrence matrices, one for each value of (d, θ). An element (i, j) of this matrix is a count of the number of pixel pairs with intensities i and j, for the given d and θ values. Various features are computed from these co-occurence matrices.

These measures were successfully used for classification of different textures such as wood, corn, grass and water (using four co-occurrence matrices with $d = 1$ and $\theta = 0$, 45, 90 and 135 degrees). However, this technique has many shortcomings. Firstly, if the number of grey levels is large, say 256, the co-occurrence matrix has 256 rows and 256 columns, and a large region is required for useful estimation of the statistics. The number of grey levels can be reduced by compressing the range, but this may introduce texture artifacts. Sometimes, only the difference of grey levels is used in computing the co-occurrence matrices. It is also necessary to limit the number of values of d and θ, and methods for automatic choice of these parameters are unclear.

e) Other Features

Faugeras and Pratt have described a different approach to estimation of texture statistics [Faugeras 1980]. Their model assumes that underlying texture is generated as an independent, identically distributed array, say $W(j, k)$, and the observed texture, say $F(j, k)$ is obtained by a spatial operator, applied to the array $W(j, k)$. It is possible to estimate $W(j, k)$ from observed $F(j, k)$, by a so-called whitening transformation. The optimal linear whitening transformation can be estimated from the auto-correlation function of the observed texture. Pratt and Faugeras also demonstrated that auto-correlation alone is not sufficient for human texture discrimination and that the Julesz conjecture holds for correlated as well as uncorrelated texture fields. In their system, final texture features are derived from measurements on the decorrelated array and from the auto-correlation function.

Davis, Johns and Aggarwal have introduced a concept of Generalized Co-occurrence Matrices [Davis 1979b]. These matrices measure the co-occurrence of some selected features in the image, and the co-occurrence is defined by satisfying a selected predicate relationship between the features. Thus, if the selected features are grey levels and the co-occurrence property is that of equality, we get the usual grey level co-occurrence matrices. However, if the selected features are binary edges and the co-occurrence predicate requires certain angular relations between the edges (such as equality or orthogonality), other properties of texture are measured. Measures similar to those used by Haralick et al. for grey level co-occurrence matrices are used for the generalized co-occurrence matrices.

A variety of other measures have been suggested. An excellent survey can be found in [Haralick 1979]. Use of Markov models or time series analysis is described in [McCormick 1974, Garber 1981]. Other measures using estimation theory methods and models for random placement of elements may be found in [Deguchi 1978], [McCormick 1975], [Modestino 1980], [Schacter 1979].

3.2.2 – Structural Texture Descriptions

The structural approaches to texture attempt to isolate the primitives that form a texture and describe the relations between these primitives in the texture pattern. However, structural descriptions are difficult to compute for natural textures, as frequently neither the primitives nor their pattern are completely uniform and regular. Further, texture patterns may be hierarchical, i.e., a particular texture pattern may repeat to form a large texture pattern and so on. Structural texture description techniques were suggested in early work e.g., see [Pickett 1970, Hawkins 1970], but not implemented until recently.

The structural view of texture has similarities with viewing sentences in a language as consisting of certain primitives related by allowed rules of a grammar. Zucker has postulated that natural textures may be viewed as being generated from a two step process [Zucker 1976]. In the first step, structured patterns, e.g., a rectangular grid, are generated according to certain rules. In the next step, these patterns undergo a transformation that introduces irregularity, in either a deterministic or a random way, to yield natural textures. However, such models are only partially successful in simulating natural textures.

An important step in generating structural descriptions is to find the appropriate primitives of a given texture. These primitives can be expected to correspond to physical objects, however, choice of such primitives requires ability of good segmentation at a detailed level. Since such capabilities do not exist in current systems for complex images, simpler primitives have been used.

Several researchers have used regions of uniform, or near uniform, intensity as primitive texture elements [Maleson 1977, Tomita 1979, Nagao 1980]. These elements are computed by the usual region segmentation techniques described earlier. Descriptions of texture elements consist of region properties such as intensity, size, shape and direction of elongation [Maleson 1977, Tomita 1979]. Some textures may have random distribution of these element properties, whereas others have elements uniform in one or more of these properties. Further distinctions between textures are based on the relations between the primitives. In work of Maleson et al [Maleson 1977], the relations used were the collinearity and parallelism of the axes of ellipses approximating the regions. Others have used statistics of "relative vectors" between texture elements [Nagao 1980, Davis 1979a]. A relative vector is given by the relative coordinates of the line joining the centroids of two textures. These statistics can also be used for syntehsis of the textures. Nagao and Matsuyama were also able to describe artificial, regular textures in a hierarchical pattern.

Davis has presented a technique for describing texture pattern of dot textures. The dot patterns themselves may represent objects of interest in a real image; in the example given in this work, they are at the center of the trees in an aerial orchard image [Davis 1979a]. He used peaks in the histogram of the directions of line joining a point to its k nearest neighbors to detect regularity of textures; a square pattern, for example, should have to peaks separated by 90 degrees.

A technique that does not require successful segmentation first is to use line segments or more simply, the local edges. If a texture is periodic, we can expect the boundaries of primitives and the local edges to occur at periodic intervals. Some simple properties of textures can be inferred by computing "edge co-occurrence" measures, defined to be the number of edges that occur at a distance d, in a given direction θ. The edges contributing to the matrix are required to be normal to the direction θ.

Vilnrotter, Nevatia and Price have used such plots for analysis of natural scenes, and their technique seems to be useful for distinguishing between highly periodic, random, and semi-periodic textures [Vilnrotter 1980, Vilnrotter 1981]. For periodic natural textures, this technique finds the perioc' and the width of the elements, in one or more directions, using the edge co-occurrence plots described above. For non-periodic textures, a primitive element width can still be found for some textures such as grass, water and sand. The width of the primitives is then used to actually isolate the primitives in the edge image, by searching for edges of opposite contrast separated by the known width. Other properties of primitives, such as length and area can be computed now. Descriptions of the primitives and their arrangements are used for recognition of textures and the recognition accuracies are claimed to be quite high, with confusion between similar textures differing in detail. These descriptions are also used for synthesis of regular, homogeneous textures.

Measures for inferring properties of textures from edge analysis that correspond to human observations are also described in [Tamura 1978]. Marr has suggested using elements of the primal sketch for texture description but has not specified the grouping properties to be used [Marr 1976].

Some of the structural properties can also be inferred from an analysis of the grey level co-occurrence matrices, see [Conners 1980a].

3.2.3 – Comparison of Texture Features

Comparison of various texture features is complicated by the number of parameters to be considered. A large number of texture types are possible. Moreover, different samples of similar objects, e.g., grass, fields, may have similar but different textures. Ideally, texture features should be invariant to changes within a class, but different from other texture classes. Due to these difficulties and the large number of suggested techniques no authoritative comparisons have been reported.

Weszka and Rosenfeld have reported a limited study and conclude that the measures based on co-occurrence matrices are better than Fourier features [Weszka 1976]. Laws claims performance suprior to that of co-occurrence features [Laws 1980]. However, these conclusions are based on a limited set of textures used in testing.

Conners and Harlow give a theoretical analysis of the performance of some texture measures for certain types of textures [Conners 1980a]. Not surprisingly, the second order measures are concluded to be superior.

Zobrist and Thompson have studied the use of a linearly weighted combination of a number of texture features [Zobrist 1975]. The weights were determined to agree with human judgements about the degree of dissimiliarities between given textures on a training set. The features with the highest weight were concluded to be the simplest three Haralick features (among the limited set that was tried).

3.2.4 – Texture Segmentation

One or more of the texture descriptors, constituting a feature vector, can be used to classify regions into one of the known types. These features can also be used for segmentation by edge or region techniques, analogous to other multi-dimensional features, such as color. Given two texture feature vectors, we need to decide if they both belong to the same surface. However, texture is not a property of a single pixel, but of a region around it. Thus, if texture is measured by centering a window of a fixed size around each pixel, this window will encompass more than one texture near the boundary. Such techniques may lead to poorly defined boundaries, with possible uncertainty in position equal to half of the window size. Another difficulty is in the choice of appropriate window size.

A model of variation in measured texture properties near the edge would be helpful in determining the appropriate window size and precise edge location (see [Davis 1980]). Due to the lack of suitable models, the common approach is to assume that the measured texture properties change smoothly from values for one texture to that of another. In this case, the edge detection corresponds to detecting a smooth ramp edge in an intensity image. Using an operator that measures gradient and choosing the gradient peak (i.e., non-maximal suppression), as described in Chapter 7, should lead to correct edge location. Some experiments are described in [Davis 1980, Thompson 1977].

The major application of texture analysis has been for images taken from airplanes or satellites. Such images usually have large, highly textured areas, e.g. forests and mountains. A common use is for classification of agricultural crops from LANDSAT satellite images.

3.3 GROUPING OPERATIONS

3.3.1 – Introduction

Most research on image segmentation has been done at the level of edge detection or region-based texture descriptions. However, there are many higher-level relationships between image features which could be used in the process of performing stereo correlation or image interpretation. Examples of this type of grouping include the detection of colinearity, curvilinearity, parallelism, bilateral or rotational symmetry, proximity, and connectedness. This higher-level grouping and segmentation can still be carried out in a bottom up manner with no specific knowledge of the objects in the image.

From groupings such as these it is possible to make monocular interpretations of three-space relations as described in [Binford 1981] and [Lowe 1981]. These interpretations and groupings can reduce ambiguity in performing object recognition or stereo correlation in much the way that edge detection reduces search and ambiguity as compared with simple intensity data. In the case of stereo correlation, to the extent that these groupings reflect meaningful three-space relations, they are likely to be present in both images and to have many characteristics in common which can be used to correlate them.

These grouping operations are all based on the detection of statistically meaningful alignments of features in the image which are unlikely to occur by accident. For example, if colinearity is detected between oriented image features, it is likely that these features are also colinear in three-space; the only other alternative is that the camera has been coincidentally placed in a restricted plane to produce the alignment. Assuming that the features are colinear in three-space provides valuable information for forming an interpretation, and also tells us that the features will be colinear when viewed from another viewpoint (eg., in the second image of a stereo pair).

Comparatively little research has been performed on the problems of forming these higher-level groupings. [Marr 1976] describes how the *full primal sketch* should contain various groupings of the data provided in the *raw primal sketch,* including curvilinear aggregation and various statistical measures of orientation and density. While Marr provides important and persuasive arguments for the importance of forming these groupings, he does not specify exactly what should be computed or how the computations can be carried out. Various other authors have looked at specific types of grouping: The Hough transform [Hough 1962], [Duda 1973] has been used to detect collinearity; [Kanade 1981] has examined the use (but not the detection) of skewed bilateral symmetry; and many authors have examined the problems of detecting "corners" [see other section of this survey]. The problem of forming higher-level segmentations is a general case of the problem of structural texture descriptions, and various researchers have examined the problems of describing regular textures [Julesz 1975], [Schatz 1977], see also survey section on structural texture descriptions].

There is a considerable amount of psychological literature on the topic of grouping and segmentation of image features. This was a favorite topic of study for Gestalt psychologists earlier in this century [Wertheimer 1923]. The Gestalt psychologists were trying to show that visual forms are perceived as a "whole" rather than as a collection of features, and therefore experimented with many ways in which visual forms are spontatneously grouped together. They noted that image elements were grouped on the basis of proximity, colinearity, continuity, and "similarity," and tried to form laws to explain the formation of figure/ground distinctions. Although their explanations (often given in terms of "attractive forces between similar elements") are of little use to computational vision, the body of experimental data provides valuable information on the grouping capabilities of the human visual system. Recent psychological research [Julesz 1975] has concentrated on the grouping and discrimination of regular textures, which shows that human vision spontaneously detects only certain types of image regularities. An important result from these psychological studies is that human vision spontaneously (in less than 200ms, before the eye can make even one saccade) performs a wide class of image groupings over the entire range of the visual field. This is a strong argument for the importance of higher-level grouping and segmentation modules for a general purpose computer vision system.

3.3.2 – Julesz 1975

Julesz,B., "Experiments in the Visual Perception of Texture"; Scientific American, April 1975, pp 34-43.

Discusses which textures can be discriminated by pure perception (ie. in about 200ms, before there is time for the eye to fix on more than one spot).

Julesz' basic theory is that textures can be discriminated if and only if there is a difference in the first or second order statistics. Since this article was written the theory has been more or less disproved. Julesz only has a couple of ways of generating textures with identical second order statistics and different third order statistics, and even on these examples the theory has some counterexamples.

First order statistics refers to the average intensity over a region. Second order statistics refers to the average intensity along each orientation in an image at different resolutions. Therefore mirror images of randomly rotated texture elements have the same second order statistics as the non-mirrored elements. Similarly, mirror images of objects symmetrical about the axis perpendicular to the mirror axis have the same second order statistics.

3.3.3 – Lettvin 1976

Lettvin, J. Y., On Seeing Sidelong, The Sciences, July 1976.

Lettvin points out many interesting experiments with peripheral vision which show that peripheral vision is limited not so much by acuity as by the ability to assemble shape in the presence of texture. For example, stare at the * in the following figure with one eye:

MOANED * N

With your gaze fixed on the *, it is easy to see the isolated N, but impossible to see the N in the word MOANED. Lettvin argues that peripheral vision does not see form so much as texture, and that therefore the surrounding texture in the word MOANED does not allow us to separate out the N. He gives other experiments to show that it is not the fact that the N is too close to the A or E which is causing the problem, so much as the similarity in texture to the A and E.

Lettvin gives many other examples where peripheral vision can see textures with high acuity, but cannot make out the individual shapes-within a textured region.

3.3.4 – Marr 1975

Marr,D., "Early Processing of Visual Information" MIT AI Memo 340 (December 1975). (replaces MIT AI Memos 324 and 334)

The paper is divided into two parts: the first part describes the computation and representation of intensity changes in the image, and the second part describes texture and grouping operations. The following comments refer only to this second part.

Marr presents this theory of texture vision as an alternative to [Julesz 1975]. Marr's theory is that "place tokens" are created for each blob, line termination and other interesting feature, and that these place tokens are grouped using curvilinear aggregation or various statistical measures. Marr claims that texture recognition is based on first-order differences computed on the edges rather than second order differences computed on intensity, as Julesz claims.

Curvilinear aggregation is the assembly of place tokens that contain an orientation into a group which preserves the orientation (eg., connecting colinear line segments). Theta-aggregation is the grouping of similarly oriented lines in a direction different from their intrinsic orientation (ie., grouping lines based on their parallelism and projecting them in a certain direction). Marr also talks about grouping into neighborhoods and regions based upon local statistical measures of density, orientation, size and contrast.

The major failure in this paper is that very little is said about how to implement these algorithms, and many aspects of the computation are left unspecified. Marr gives some examples of computations performed on real images, but no description is given of the algorithms which computed these examples and the definite impression left is that these were ad hoc programs made for specific examples. Some of the processes, such as region grouping based on statistical measures, are merely hinted at without any specification of how the statistics are used.

In summary, this paper is valuable for its overview of various types of grouping operations which are apparently done in human vision, and its rough specification of the way some of these grouping operations are carried out. However, the specific algorithms are very poorly specified, and rather little psychological evidence is brought to bear in support of the specific operations chosen. Some very hard problems are glossed over without even acknowledging their presence. Some important grouping criteria, such as symmetry, cotermination, closure, etc., are not even mentioned.

3.3.5 – Schatz 1977

Schatz,B.R., The Computation of Immediate Texture Discrimination, MIT AI Memo 426, August 1977.

This is a study of computer algorithms for doing the sort of texture discriminations which were studied by Julesz (ie., regular textures in which one rectangular region has elements different from the surrounding field). Unfortunately, Schatz did not implement any of his proposed algorithms, and there are many aspects of them that are unspecified.

Schatz claims that we discriminate the given class of textures by looking only at the length and orientation of lines, including virtual lines which are constructed between feature points. This theory is given a fair amount of support in the many computer generated textures shown in the paper. Schatz relates this theory to the one offered by Julesz, and gives convincing arguments to show that it is more tenable than Julesz' theory that human vision discriminates textures iff the second order statistics are different.

3.3.6 – Stevens 1976

Stevens,K.A., Occlusion Clues and Subjective Contours,
MIT AI Memo 363, June 1976.

A patient with damage to the left medial occipital region and the posterior hemispheres was unable to see subjective contours or notice monocular occlusion cues.

The patient was shown various drawings in which normal subjects would see subjective contours, or would group objects together into larger units. This patient would eventually notice some of the alignments in these images, but would apparently not do so as a basic perceptual operation. He would not arrive at the interpretation that one object in the image was occluding another, even though other aspects of his visual performance seemed relatively normal.

The main result seems to be that the detection of occlusion cues is probably related to the phenomenon of subjective contours. Grouping operations which are a part of the detection and perception of subjective contours seem to form a separate module of the visual system.

3.3.7 – Stevens 1979

Stevens,K.A., Constraints on the Visual Interpretation of
Surface Contours, MIT AI Memo 522, March 1979.

"Surface contours" are the images of curves which lie on some physical surface (ie., essentially all curves in the image of a 3D scene). Stevens outlines various constraints which could be used to recover the 3D shape of surfaces from the 2D shape of the image contours.

This paper points out various types of constraints which are probably used by the human visual system at some level of analysis. However, Stevens seems to mix together constraints based on very general assumptions with those based on very specific assumptions. For example some constraints are based on the assumption of general camera position, while others are based on the claim that "shadow casting edges are usually vertical (eg., tree trunks, ...)." No general mechanism is suggested for evaluating the correctness of the assumptions or dealing with cases in which the assumptions turn out to be wrong. It is therefore not surprising that none of these constraints were implemented or tested on real images.

However, if we ignore the questions of implementation and integration, the paper is useful as a grab-bag of ideas for constraints on the interpretation of image contours. Here is a brief list of some of them:

1) Assuming general camera position:

a) straight image curve implies straight three-space curve.

b) smooth image curve implies smooth three-space curve.

2) Assuming that skewed symmetry in the image is the result of the projection of a bilaterally symmetric object in 3-space:
We can calculate constraints on the tilt of the plane of symmetry from the degree of skew.

3) Assuming constant curvature:
Changes in curvature can be used to constrain tilt (eg. circle vs. ellipse).

4) Assuming orthographic projection and general camera position:
Parallelism in image implies parallelism in 3-space.

5) Assuming that the angle between the surface contour direction of curvature and the surface is constant (which implies that the surface is cylindrical):
Then we know the curve is either geodesic or asymptotic, which implies some constraints on the surface position under the contour.

The paper ends with a weak section arguing for the "validity" of the various constraints (ie, are the above assumptions likely to be true?). The answer is that sometimes they will be true and other times false, depending on the particular scene being viewed. No way is suggested in which the constraints could be used in the absence of a priori evidence for the truth of the assumptions.

3.4 References

[Abdou 1978] Abdou,I.E., "Quantitative Methods of Edge Detection," Ph.D. thesis, University of Southern California, July 1978. Also USCIPI Report 830. (cited on p. 92,93,116,117)

[Altes] Altes, R.A., "Spline-like Image Analysis with a Complexity Constraint. Similarities to Human Vision," unpublished paper, ca. 1975, 36 pp. (cited on p. 93,94)

[Bajcsy 1973] Bajcsy, R., "Computer Identification of Visual Surface," *Computer Graphics and Image Processing*, vol. 2, 1973, 118-130. (cited on p. 147)

[Ballard 1976] Ballard, D.H. and J. Sklansky, "A ladder-structured decision tree for recognizing tumors in chest radiographs," *IEEE Trans. Computers*, vol. C-25, 5, May 1976, 503-513. (cited on p. 94)

[Beaudet 1978] Beaudet, P.R., "Rotationally invariant image operators," in*Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, (Kyoto, Japan, Nov. 7-10, 1978), 579-583. (cited on p. 94,94,108,108,108,115,115,115,115)

[Binford 1970] Binford, T.O., "The TOPOLOGIST," Internal Report MIT-AI, 1970. (cited on p. 95)

[Binford 1981] Binford, T.O., "Inferring Surfaces from Images," *Artificial Intelligence*, 17, 1981, 205-244. (cited on p. 93,151)

[Brice 1970] Brice, C.R. and C.L. Fennema, "Scene Analysis Using Regions," Artificial Intelligence Group Technical Note 17, Stanford Research Institute, April 1970. (cited on p. 95)

[Caelli 1978b] Caelli, T., B. Julesz and E. Gilbert, "On Perceptual Analyzers Underlying Visual Texture Discrimination: Part II," *Biological Cybernetics*, vol. 29, no. 4, 1978, 201-214. (cited on p. 146)

[Chen 1980] Chen, P.C. and T. Pavlidis, "Image Segmentation as an Estimation Problem," Computer Graphics and Image Processing, February 1980, vol. 12, no. 2, 153-172. (cited on p. 96)

[Conners 1980a] Conners, R.W. and C.A. Harlow, "A Theoretical Comparison of Texture Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 1980, 204-222. (cited on p. 149,150)

[Cooper 1980] Cooper, D.B., H. Elliott, F. Cohen, L. Reiss, and P. Symosek, "Stochastic Boundary Estimation and Object Recognition," *Computer Graphics and Image Processing*, vol.12, 1980, p. 326. (cited on p. 96)

[Davis 1973] Davis, L., "A Survey of Edge Detection Techniques", TR-273, Univ of Md, Computer Science Center, 1973. (cited on p. 117)

[Davis 1979a] Davis, L.S., "Computing the Spatial Structure of Cellular Textures," *Computer Graphics and Image Processing*, vol. 11, no. 2, October 1979, 111-122. (cited on p. 149,149)

[Davis 1979b] Davis, L.S., S.A. Johns and J.K. Aggarwal, "Texture Analysis Using Generalized Co-occurrence Matrices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 3, July 1979, 251-259. (cited on p. 148)

[Davis 1980] Davis, L.S. and A. Mitiche, "Edge Detection in Textures," *Computer Graphics and Image Processing*, vol. 12, no. 1, Jan 1980, 25-39. (cited on p. 150,150)

[DeBoor 1978] DeBoor, C., **A Practical Guide to Splines**, Springer, 1978 (Vol 27 in Applied Mathematical Sciences series). (cited on p. 113)

[Deguchi 1978] Deguchi, K. and I. Morishita, "Texture Characterization and Texture-based Image Partitioning Using Two-dimensional Linear Estimation Techniques," *IEEE Transactions on Electronic Computers*, vol. 27, no. 8, Aug. 1978, 739-745. (cited on p. 148)

[Dineen 1955] Dineen, G.P., "Programming Pattern Recognition," *Proc. WJCC*, 94-100, March 1955. (cited on p. 89)

[do Carmo 1976] do Carmo, M.P., **Differential Geometry of Curves and Surfaces**, Prentice-Hall, Englewood Cliffs, N.J., 1976. (cited on p. 102)

[Dreschler 1981a] Dreschler, L. and H.-H. Nagel, "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," Report IfI-HH-M-90/81, Fachbereich Informatik, Universität Hamburg. (cited on p. 94)

[Dreschler 1981b] Dreschler, L. and H.-H. Nagel, "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (IJCAI-81), Aug 1981, Vancouver. (cited on p. 94)

[Duda 1971] Duda, R.O. and P.E. Hart, "A generalized Hough transformation for detecting lines in pictures," SRI AI Group Tech Note 36, 1971. (cited on p. 94)

[Duda 1972] Duda, R.O. and P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM* 15, no. 1, 1972, 11-15. (cited on p. 94)

[Duda 1973] Duda, R.O. and P.E. Hart, **Pattern Classification and Scene Analysis**, Wiley, New York, 1973. (cited on p. 92,92,94,96,120,151)

[Faugeras 1980] Faugeras, O.D. and W.K. Pratt, "Decorrelation Methods of Texture Feature Extraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, July 1980, 323-332. (cited on p. 147)

[Fennema 1970] Fennema, C.L. and C.R. Brice, "Scene analysis of pictures using regions", *Artificial Intelligence Journal* 1, 1970, 205-226. (cited on p. 95)

[Gagalowicz 1980] Gagalowicz, A., "Visual Discrimination of Stochastic Texture Fields based upon their second Order Statistics," *Proceedings of Fifth International Conference on Pattern Recognition*, Miami Beach, Fl., December 1980, 786-788. (cited on p. 146)

[Garber 1981] Garber, D., "Computational Models for Texture Analysis and Texture Synthesis," University of Southern California, USCIPI Report 1000, May 1981, (Ph.D. Thesis). (cited on p. 148)

[Griffith 1973] Griffith, A.K., "Mathematical Models for Automatic Line Detection," *Journal of the ACM*, vol. 20, no. 1, January 1973, p. 62. (cited on p. 96)

[Habibi 1972] Habibi, A., "Two Dimensional Bayesian Estimate of Images", *Proceedings of the IEEE*, vol. 60, no. 7, 1972, 878–883. (cited on p. 121)

[Halmos 1957] Halmos, P.R., **Introduction to Hilbert Space and the Theory of Spectral Multiplicity**, Chelsea, 1957. (cited on p. 93)

[Halmos 1963] Halmos, P.R., "What Does the Spectral Theorem Say?," *The American Mathematical Monthly*, March 1963, 241–247. (cited on p. 93)

[Haralick 1973] Haralick, R.M., K. Shanmugam and I. Dinstein, "Texture Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, Nov. 73, 610–621. (cited on p. 147)

[Haralick 1979] Haralick, R.M., "Statistical and Structural Approaches to Texture," *Proceedings of IEEE*, vol. 67, no. 5, May 1979, 786–804. (cited on p. 148)

[Haralick 1980] Haralick, R. M., "Edge and Region Analysis for Digital Image Data," *Computer Graphics and Image Processing*, vol. 12, no. 1, January 1980, 60–73. (cited on p. 94,98,132,132,133)

[Haralick 1981] Haralick, R. M., "The Digital Edge," *Proc. IEEE Conf. Pattern Recognition and Image Processing*, August 1981, 285–291. (cited on p. 94)

[Hawkins 1970] Hawkins, J.K., "Texture Properties for Pattern Recognition," in **Picture Processing and Psychopictorics**, B.S. Lipkin and A. Rosenfeld, (Editors), Academic Press, New York, 1970, 347–370. (cited on p. 148)

[Herskovits 1970] Herskovits, A. and T.O. Binford, "On Boundary Detection," MIT Project MAC, Artificial Intelligence Memo 183, July 1970. (cited on p. 91)

[Horn 1972] Horn, B.K.P., "The Binford-Horn Edge Finder," MIT AI Memo 285, 1972, revised December 1973. (cited on p. 95)

[Hough 1962] Hough, P.V.C., "Method and Means for recognizing complex patterns," U.S.Patent 3,069,654, December 18, 1962. (cited on p. 94,151)

[Hsu 1978] Hsu, S., J.L. Mundy, P.R. Beaudet, "Web Representation of Image Data," *Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, (Kyoto, Japan, Nov. 7–10, 1978), 579–583. (cited on p. 94)

[Hueckel 1969] Hueckel, M.H., "An Operator which Locates Edges in Digital Pictures," Stanford Computer Science Dept. Memo AIM-105, Oct. 1969. (cited on p. 93)

[Hueckel 1971] Hueckel, M.H., "An Operator which Locates Edges in Digital Pictures," *JACM*, vol. 18, no. 1, January 1971, 113–125. Erratum in 21, 1974,350. (cited on p. 93,127,128)

[Hueckel 1973] Hueckel, M.H., 'A Local Visual Operator Which Recognizes Edges and Lines," *JACM*, vol. 20, no. 4, October 1973, 634–647. (cited on p. 127,128)

[Julesz 1962] Julesz, B., "Visual Pattern Discrimination," *IRE Transactions on Information Theory*, vol. 8, February 1962, 84–92. (cited on p. 146)

[Julesz 1975] Julesz, B., "Experiments in the Visual Perception of Texture," *Scientific American*, Apr. 1975, 34–43. (cited on p. 151,152,153)

[Julesz 1978] Julesz, B., E.N.Gilbert and J.D. Victor, "Visual Discrimination of Textures with Identical Third-Order Statistics," *Biological Cybernetics*, vol. 31, no. 3, 1978, 137-140. (cited on p. 146)

[Kanade 1981] Kanade, T., "Recovery of the Three-Dimensional Shape of an Object from a Single View," *Artificial Intelligence*, vol. 17, 409, 1981. (cited on p. 151)

[Kirsch 1971] Kirsch, R.A., "Computer Determination of the Constituent Structure of Biological Images," *Computers and Biomedical Research*, vol. 4, no. 3, 1971, 315–328. (cited on p. 92,95)

[Landau 1961] Landau, H.J. and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — II," *Bell Syst. Tech. J.*, 40, January 1961, 65–84. (cited on p. 122,138)

[Landau 1962] Landau, H.J. and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — III: The Dimension of the Space of Essentially Time- and Band-Limited Signals," *Bell Syst. Tech. J.*, 41, July 1962, 1295–1336. (cited on p. 122,138)

[Laws 1980] Laws, K.I., "Textured Image Segmentation," University of Southern California Report USCIPI 940 (Ph.D. thesis), Jan. 1980. (cited on p. 147,149)

[Lowe 1981] Lowe, D.G., and T.O. Binford, "The Interpretation of Geometric Structure from Image Boundaries," *Proc. ARPA Image Understanding Workshop*, 39–46, April 1981. (cited on p. 151)

[Machuca 1981] Machuca, R. and A.L. Gilbert, "Finding edges in noisy scenes," *Pattern Analysis and Machine Intelligence*, PAMI-3, no. 1, January 1981, p. 303. (cited on p. 96)

[Macrenhas 1978] Macrenhas, N.D.A. and L.O.C.Prado, "Edge detection in images: a hypothesis testing approach," in *Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, Kyoto, Japan, Nov.7-10, 1978. (cited on p. 96)

[Maleson 1977] Maleson, J.T., C.M. Brown and J.A. Feldman, "Understanding Natural Texture," *Proceedings of ARPA Image Understanding Workshop*, Palo Alto, Ca., October 1977, 19–27. (cited on p. 148,148,149)

[Marr 1976] Marr, D., "Early Processing of Visual Information," *Philosophical Transactions of Royal Society of London*, B275, 1976, 483–524. (cited on p. 149,151)

[Marr 1979] Marr, D. and E. Hildreth, "Theory of Edge Detection," AI Memo 518, MIT AI Lab, April 1979. Also Proc.R.Soc.Lond.B., 1980, 207, 187-217. (cited on p. 94,138)

[Martelli 1972] Martelli, A., "Edge Detection using Heuristic Search Methods," Dept of EE and Computer Science, NYU, University Heights, Bronx, NY, 10453. Also *Computer Graphics and Image Processing*, 1, 1972, 169–182. (cited on p. 91,95)

[Martelli 1973] Martelli, A., "An Application of Heuristic Search Methods to Edge and Contour Detection," Instituto di Elaborazione della Informazione del Consiglio Nazionale delle Richerche, Pisa, 1973. Also *Comm. ACM* 19, 1976, 73–83. (cited on p. 91,95)

[McCormick 1974] McCormick, B.H. and S.N. Jayaramamurthy, "Time Series Model for Texture Synthesis," *International Journal of Computer and Information Science*, vol. 3, no. 4, 1974, 329–343. (cited on p. 148)

[McCormick 1975] McCormick, B.H. and S.N. Jayaramamurthy, "A Decision Theory Method for the Analysis of Texture," *International Journal of Computer and Information Science*, vol. 4, no. 1, 1975, 1-37. (cited on p. 148)

[Modestino 1980] Modestino, J.W., R.W. Fries and A.L. Vickers, "Stochastic Image Models Generated by Random Tesselations of the Plane," *Computer Graphics and Image Processing*, vol. 12, 1980, 74-98. (cited on p. 148)

[Montanari 1970] Montanari, U., "On the Optimal Detection of Curves in Noisy Pictures," Artificial Intelligence Laboratory, Stanford University, Memo AIM-115, 1970. (cited on p. 95,124)

[Montanari 1971] Montanari, U., "On the Optimal Detection of Curves in Noisy Pictures," *Comm. ACM* 14, May 1971, 335-345. (cited on p. 95,124)

[Morse 1953] Morse, P.M. and H. Feshbach, **Methods of Theoretical Physics, Part II**, McGraw-Hill, 1953. (cited on p. 116)

[Nagao 1980] Nagao, M. and T. Matsuyama, **A structural Analysis of Complex Aerial Photographs**, Plenum Press, New York, 1980. (cited on p. 148,149)

[Nevatia 1978] Nevatia, R. and K.R. Babu, "Linear Feature Extraction," *Proc. ARPA Image Understanding Workshop*, Pittsburgh, November 1978, 73-78. (cited on p. 95)

[O'Gorman 1976] O'Gorman, F., "Edge Detection using Walsh Functions," *Proc AISB*, p 195, July 1976. Also: *Artificial Intelligence* 10, 1978, 215-233. (cited on p. 93,131)

[Ohlander 1975] Ohlander, R.B., "Analysis of Natural Scenes," Dept of Computer Science, Carnegie-Mellon Univ, April 1975. (PhD thesis) (cited on p. 95)

[Pavlidis 1972] Pavlidis, T., "Segmentation of Pictures and Maps through Functional Approximation," *Computer Graphics and Image Processing*, vol. 1, 1972, 360-372. (cited on p. 118)

[Pavlidis 1977] Pavlidis,T., **Structural Pattern Recognition**, Springer-Verlag, 1977. (cited on p. 96)

[Pickett 1970] Pickett, R.M., "Visual Analysis of Texture in the Detection and Recognition of Objects," in *Picture Processing and Psychopictorics*, B.S. Lipkin and A. Rosenfeld, (Editors), Academic Press, New York, 1970, 289-308. (cited on p. 148)

[Prewitt 1970] Prewitt, J.M.S., "Object Enhancement and Extraction," in **Picture Processing and Psychopictorics**, B.S.Lipkin and A.Rosenfeld,Eds., Academic Press, New York, 1970. (cited on p. 94,112)

[Purks 1977] Purks, S.R. and W. Richards, "Visual Texture Discrimination Using Random Dot Patterns," *Journal of Optical Society of America*, vol. 67, June 1977, 765-771. (cited on p. 146)

[Roberts 1963] Roberts, L.G., "Machine Perception of Three-Dimensional Solids," in **Optical and Electro-Optical Information Processing, Optical and Electro-Optical Information Processing**, J.T.Tippett et al., Eds., MIT Press, Cambridge, Mass., 1965, 159-197. Also Technical Report no. 315, Lincoln Laboratory, MIT (May 1963). (cited on p. 95,111,112)

[Rosenfeld 1975] Rosenfeld, A., R.A. Hummel, S.W. Zucker, "Scene Labelling by Relaxation Operations," Computer Science Center, Univ of Md, TR-379, May 1975. Also

IEEE Trans. Syst. Man Cybern., SMC-6, no. 6, June 1976, 420–433. (cited on p. 95)

[Rosenfeld 1976] Rosenfeld, A. and A.C. Kak, **Digital Picture Processing**, Academic Press, New York, 1976. (cited on p. 92)

[Rutkowski 1978] Rutkowski, W.S. and A. Rosenfeld, "A Comparison of Corner-Detection Techniques for Chain-Coded Curves," University of Maryland Technical Report TR-623, Jan. 1978. (cited on p. 96)

[Santalo 1976] Santalo Sors, L.A., "Integral Geometry and Geometric Probability," Addison-Wesley, 1976 (Vol 1 in Encyclopedia of Mathematics and its Applications). (cited on p. 114)

[Schacter 1979] Schachter, B. and N. Ahuja, "Random Pattern Generation Process," *Computer Graphics and Image Processing*, vol. 10, 1979, 95–114. (cited on p. 148)

[Schatz 1977] Schatz, B.R., "The Computation of Immediate Texture Discrimination," MIT AI Memo 426, August 1977. (cited on p. 151)

[Shafer 1980] Shafer, S.A., "MOOSE. Users' Manual, Implementation Guide, Evaluation," Bericht 70, IfI-HH-B-70/80, Fachbereich Informatik, Universität Hamburg, April 1980. (cited on p. 95)

[Shanmugam 1979]Shanmugam, K.S., F.M. Dickey, J.A. Green, "An optimal frequency domain filter for edge detection in digital images," *IEEE Trans Pattern Analysis and Machine Intelligence*, PAMI-1, Jan. 1979, 39–47. (cited on p. 92,122)

[Shapiro 1974] Shapiro, S.D., "Detection of lines in noisy pictures using clustering," *Proceedings of the Second International Joint Conference on Pattern Recognition*, Copenhagen, Aug. 13–15, 1974, 317–318. (cited on p. 94)

[Shapiro 1975] Shapiro, S.D., "Transformations for the Computer Detection of Curves in Noisy Pictures," *Computer Graphics and Image Processing*, 4, 1975, p. 328. (cited on p. 94)

[Shapiro 1978] Shapiro, S.D., "Generalization of the Hough Transform for Curve Detection in Noisy Digital Images," *Proceedings of the Fourth International Joint Conference on Pattern Recognition* (IJCPR-78), 710–714. (cited on p. 94)

[Shaw 1977] Shaw, G.B.,"Local and Regional Edge Detectors: Some Comparisons," Univ. of Maryland Technical Report TR-614, December 1977. (cited on p. 117)

[Shaw 1979] Shaw, G.B.,"Local and Regional Edge Detectors: Some Comparisons," Computer Graphics and Image Processing, vol. 9, no. 2, Feb. 1979, 135–149. (cited on p. 117)

[Shirai 1975] Shirai, Y., "Edge finding, segmentation of edges and recognition of complex objects," Proc. 4th IJCAI, 1975, 674-681. (cited on p. 96)

[Slepian 1961] Slepian, D. and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — I," *Bell Syst. Tech. J.*, 40, January 1961, 43–63. (cited on p. 122,138)

[Somerville 1976] Somerville, C. and J.L. Mundy, "One Pass Contouring of Images Through Planar Approximation," *Proc. of the 3rd International Joint Conference on Pattern Recognition (IJCPR-76)*, Nov. 1976 (IEEE 76CH1140-3C). (cited on p. 95)

[Tamura 1978] Tamura, H., S. Mori and T. Yamawaki, "Textural Features Corresponding to Visual Perception, " *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, June 1978, 460-473. (cited on p. 149)

[Thompson 1977] Thompson, W., "Textural Boundary Analysis," *IEEE Transactions on Computers*, vol. 26, 1977, 272-276. (cited on p. 150)

[Tomita 1979] Tomita, F., Y. Shirai and S. Tsuji, "Description of Textures by a Structural Analyzer," *Proceedings of the International Joint Conference on Artificial Intelligence*, Tokyo, August 1979, 884-889. (cited on p. 148,148)

[Turner 1974] Turner, K., "Computer Perception of curved objects using a television camera," Ph.D. dissertation, Edinburgh University, November 1974. (cited on p. 91)

[Victor 1978] Victor, J.D. and S. Brodie, "Discriminable Textures with Identical Buffon Needle Statistics," *Biological Cybernetics*, vol. 31,no. 4, 1978, 231-234. (cited on p. 146)

[Vilnrotter 1980] Vilnrotter, F., R. Nevatia and K. Price, "Structural Description of Natural Textures," *Proceedings of Fifth International Pattern Recognition Conference*, Miami, Dec. 1980. (cited on p. 149)

[Vilnrotter 1981] Vilnrotter, F., "Structural Analysis of Natural Textures," University of Southern California, Ph. D. Thesis, USCISG 100, September 1981. (cited on p. 149)

[Weszka 1976] Weszka, J., C.R. Dyer and A. Rosenfeld, "A Comparative Study of Texture Measures for Terrain Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, no. 4, April 1976, 269-285. (cited on p. 149)

[Yakimovsky 1976] Yakimovsky, Y., "Boundary and Object Detection in Real World Images," *Journal of the ACM*, vol. 23, no. 4, October 1976, p. 599. (cited on p. 96)

[Zobrist 1975] Zobrist, A. and W. Thompson, "Building a Distance Function for Gestalt Grouping," *IEEE Transactions on Computers*, vol. 24, 1975, 718-728. (cited on p. 150)

[Zucker 1976] Zucker, S.W., "Toward a Model of Texture," *Computer Graphics and Image Processing*, vol. 5, 1976, 190-202. (cited on p. 148)

[Zucker 1977] Zucker, S.W., R.A. Hummel, and A. Rosenfeld, "An application of relaxation labelling to line and curve enhancement," *IEEE Trans. Computers* C-26, 1977, 394-403. (cited on p. 95)

Chapter 4

IMAGE REGISTRATION

4.1 Registration Issues

The basic problem of image registration (inversely, scene registration) is the following: given two 2-dimensional images of a 3-dimensional scene, find the spatial relationships between the cameras which produced the images. The scene may be stationary, viewed simultaneously with two distinct cameras or by one camera in motion, or the camera may be stationary with the scene or objects within it moving.

Images alone are insufficient for the determination of camera relative positions and attitudes (the ***camera solution***) — additional information or assumptions about the scene or about the cameras used in the imaging is necessary for the registration. Further information is often used in addition, to make the registration computationally easier. The primary general assumptions are those of object coherence and rigidity. In general, proximal image points correspond to proximal object points. A scene is made up of rigid objects, so that image changes not due to camera differences are due to the rotation and translation of rigid objects. These assumptions form the basis of a common approach to registration, in which points in one image are matched to points in other images thought to correspond to the same 3-D object point. The matchings of a set of such points is then used to determine the spatial relationships between the two imaging points (relative to the scene, or relative to one of the cameras). Once the camera solution is found, the determination of the object point 3-D positions can be made. The determination of the matchings is a substantial problem, and is dealt with in the Stereo Systems section and the Correspondence Constraints section of this survey.

A further problem of registration is that the data are often noisy: cameras have a limited resolution and are hindered with distortion and noise, and image point matchings may be inaccurate. The best that can then be done in these circumstances is to find an estimate of the camera relationships that minimizes some error function. The error function is often a measure of some fit in a least-squares sense. Such error function minimization generally requires the use of an iterative numerical procedure, which takes a given estimate and improves upon it through hill-climbing (or, more appropriately for a minimization, hill-falling). Important characteristics of such procedures include rate of convergence, convergence properties when starting with a poor estimate, sensitivity to wild points, and the complexity of each iteration step.

Procedures for registration can thus be characterized by the additional information they use or additional assumptions they make, and by the characteristics of the *numerical techniques* they employ. Some common classes of procedures are indicated below.

4.1.1 – The assumption of small camera differences and optic flows

It may be that the images given in a registration problem arise from views of a scene differing through incremental changes in imaging position or incremental changes in imaging time.

The resulting gradual change from image to image is termed the '*optic flow*', and is represented computationally by a vector at each pixel pointing in the direction of change, or flow. This flow information provides much the same information as a matching of image points, and can be used similarly for registration. One distinction between optic flow calculations and general matchings is that the former usually assume that the difference between images, or the difference in position between cameras recording the images, is small. These special conditions allow the use of techniques based on derivatives, where it is assumed that differences between corresponding pixels on the images approximate the temporal derivative of the image intensity at the pixel. This assumption, combined with a similar assumption about the differences between nearby pixels approximating spatial derivatives, allows facts about derivatives to be used in estimating the optic flow or solving the registration problem. [Lucas 81] describes a fairly general technique for registration using these kinds of assumptions.

4.1.2 – The assumption of camera differences of a fixed kind

In general, the spatial relationship between two cameras can be described as a *translation*, indicating the difference in location of the cameras, and a *rotation*, indicating the difference in orientation of the cameras. These relations are determined by a translation vector between the cameras, and three parameters indicating the rotation. The numerical estimation problem is easier, and the registration problem more quickly solved, if the translation or rotation are known exactly. Thus, it is often assumed, or known, that there is no rotation: the cameras have the same orientation, and the objects in the scene don't rotate. The only thing remaining to be determined is the translation. If the translation in turn is assumed to be parallel to a plane, say the picture plane, then the estimation task is even easier. Similarly, sometimes it is assumed that the translation is zero, and only the rotation need be determined.

A common technique when neither the rotation nor the translation is known is to use an estimate of the rotation to determine the translation. The 'fit' of this translation then becomes an error function of the rotation alone. [Prazdny 81a] uses this technique in the context of optic flows, while [Clarkson 81] and [Nagel 81] describe a similar technique in the context of image point matches. The basic strategy of these techniques builds upon the fact that when no rotation is present, three particular vectors are co-planar: the translation vector from the projection center of the first camera to the projection center of the second, that from the center of the first camera to a given object point, and that from the center of the second camera to the given object point. This condition is true for all object points, with each image point match confining the translation vector to a plane. Thus several such matches will determine the translation vector. In the context of optic flow calculations, this co-planarity condition implies the existence of a *focus of expansion,* or FOE. Consider where the three rays corresponding to the three vectors mentioned above intersect a superposition of the two image planes. Two of these points of intersection are the image points corresponding to the object point, and the difference between them results from the translation between the two cameras. Because all three rays are co-planar, their intersection points with the image plane are co-linear. Thus, the vector between the image points corresponding to a given object point is directed toward the point of intersection of the translation vector with the joint image plane. The set of such vectors will be seen to radiate from this point of intersection. This point is called the *focus of expansion,* for this reason. It is the focus, or center, of expansion of an image as a camera moves forward: all optic flow vectors point to (or away from) the FOE. This analysis assumes that no camera rotation is present. If the optic flow vectors do not point to an FOE, a rotation is present which has not been compensated for. The consistency of the optic flow in this sense thus implies an error condition on an estimate of the rotation which can then be numerically minimized.

Another aspect of the registration problem, related to assumptions about camera relationships, is the question of the number of objects in the scene which move independently. As mentioned, most registration systems assume that image points to be considered correspond to one rigid body which moves before a stationary camera or is viewed from different directions. Some systems handle the presence of a moving articulated (jointed) body, and a few even allow amorphous shape changes.

4.1.3 – Assumptions about the image formation process:

Perspective vs. parallel projection

The parallel projection of object points to the image plane is a common approximation to the perspective projection model of image formation. Perspective projection is a more realistic imaging model but is more difficult to use, both mathematically and computationally. The parallel model is often used for convenience, and is adequately realistic when the object points are far away from the cameras. This is the case in, for example, aerial imaging in the field of photogrammetry.

The use of control points or distance information.

Sometimes the images under consideration contain one or several ***control points***: points in the image corresponding to known object points. Sometimes the distance to an object being photographed is known. These and other kinds of 'control' information have been exploited in the field of *close-range photogrammetry* (see, for example, [Wong 75]).

4.2 Registration Summary

To summarize, some salient questions about systems for registration are:

- What classes of motion are considered:
    - 2 dimensional translations?
    - 3 dimensional translations?
    - With rotation?
    - Without rotation?
    - Large motions?
    - Small motions (allowing gradient approximations)?

- What kinds of viewed objects are considered:
    - Rigid?
    - Rigid articulated?
    - Amorphous?

- What model of image formation is used:
    - Perspective?
    - Projective?
    - Something else?

- Are control points in the images, or distance estimates available?
- Is the numerical procedure used:
    - Robust when noise and wild points are present?
    - Rapidly convergent?
    - Convergent when given a poor starting estimate?

4.3 Image Sequence Summaries

4.3.1 – Clarkson

"A Procedure for Camera Calibration," K. L. Clarkson Proceedings DARPA Image Understanding Workshop, p. 175–177, April 1981.

A procedure for the camera calibration problem using image point matches with perspective projection is described, in the context of a stationary scene with two different cameras. An error function with a linear translational component and a non-linear rotational component is described, and the numerical method of variable projection is outlined for solving the separable least-squares problem of minimizing this error. The basis of the error condition is the co-planarity condition on the translation vector and projection center — image point rays described in the introduction above. When no camera rotation is present, the rays from the centers of projection of the two cameras considered and the translation vector between the cameras are all co-planar, so that the cross-product of the two rays should be perpendicular to the translation. Since this is true for all such rays, a set of cross-product vectors can be calculated which should all be in a plane to which the translation is normal. If these vectors are not co-planar, a camera rotation is present which has not been dealt with, so that we have the error function described. The numerical procedure used is described as rapidly convergent when given a good initial estimate of the rotation, but is somewhat sensitive to the initial estimate.

4.3.2 – Clocksin

"Perception of surface slant and edge labels from optical flow: a computational approach," William F. Clocksin, Perception, 1980, Vol.9, 253-269.

This paper deals with the inference of surface slant in a static scene viewed through a moving camera. Surface edges are categorized as concave, convex, occluding, disoccluding, or contour. A contour is an edge separating a surface from the (infinitely distant) background. It is not obvious that one can distinguish these (or in fact wants too) on the basis of optical flow. The processing seems to make the separation on the fact that in contour edges, the flow drops to zero at the infinity point ... what about noise/cloud, etc? The first analysis is in terms of a given depth map (which is pretty obvious). The author then says that its true for a $\delta^2 f$ (f is optical flow, δ^2 is a laplacian). Singularities are said to indicate the edge types.

4.3.3 – Fennema and Thompson

"Velocity Determination in Scenes Containing Several Moving Objects," Claude L. Fennema and William B. Thompson, Computer Graphics and Image Processing, 9, 1979.

This paper describes a non-matching motion analysis scheme (GITM for Gradient Intensity Transform Method) applicable for rigid body translation not involving rotation. The time variation in intensity and the spatial gradient at each point place constraints on the speed and direction of scene surface points. A modified Hough transform is used to cluster surface elements to obtain a unique velocity vector for that surface. Image blurring is used to reduce noise and diminish the effect of small scale texture. Images processed were 128 by 128 with 8 bits of grey each, being assorted Snoopy/Charlie Brown toys, and a few scattered tools. Results suggest that velocity angles can be determined to within about $\frac{\pi}{8}$ and magnitude to about 1 pixel per frame.

4.3.4 – Gennery

"Modelling the Environment of an Exploring Vehicle By Means of Stereo Vision," Donald B. Gennery, Ph.D. thesis, Department of Computer Science, Stanford University, AIM-339, June 1980.

The chapter of this thesis dealing with camera calibration develops an error function for the registration problem with inputs of image point matchings. The image formation model used is perspective projection, requiring 2 views of five points on a rigid object. A minimization method based on Gauss-Newton iteration is described, with an analysis of the statistical and numerical issues involved. The method described includes the use of editing out of wild points in the data and weightings for the image point matchings.

4.3.5 – Jacobus, Chien, Selander

"Motion Detection and Analysis of Matching Graphs of Intermediate-Level Primitives," Charles J. Jacobus, Robert T. Chien, John Michael Selander, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 6, Nov 1980.

A technique for the matching of non-isomorphic graphs in the context of object recognition is discussed here. Graph entities are said to be intermediate-level structures, each a significant part of a whole object. The principal unit of the graph (referred to as the '*half-chunk*') is a segment of region boundary of constant curvature (bounded by curvature discontinuities). These half-chunks are linked together in forming object descriptions by 5 types of links - these are the obvious links to regions bounded, connecting segments, and paired half-chunks in the opposite sense. 13 feature predicates are associated with each graph node, and these are parameterized in a ternary system of <true>, <false>, and <?>(undecided). Unambiguous node matches are used as seeds to propagate confidences through ambiguous node matchings, and a "relaxation" scheme ruminates over the structure in search of a consistent total interpretation. Matched graph results may be interpreted as depth pairings (with one set of inference rules) or as motion pairings (using a looser set of rules). This research has relevance to the problem of using monocular cues to stereopsis.

4.3.6 - Lawton

"Optic flow field structure and processing image motion," Daryl T. Lawton, Seventh Int. Joint Conf. on Artificial Intelligence, Vancouver, B.C., p. 700-703, August 1981.

The use of optic flows to determine object motion is briefly considered for three restricted cases: known camera motion, translational motion only, and motion in a plane. A perspective projection model is used.

4.3.7 - Lucas and Kanade

"An Iterative Image Registration Technique with an Application to Stereo Vision," Bruce D. Lucas and Takeo Kanade, Seventh Int. Joint Conf. on Artificial Intelligence, Vancouver, B.C., p. 674-679, August 1981.

This report describes an iterative technique for adjusting estimates to camera registration parameters based on spatial intensity gradients. It uses a Newton-type iteration, with initial estimates of camera parameters, to converge on the registration parameters. The technique can be extended to registration between images related by arbitrary linear transformations such as rotation, scaling, and shearing. The paper also demonstrates the use of this spatial intensity gradient iterative matching technique for stereo correspondence. Having the camera parameters and an initial estimate for the depth of chosen points in one view, the iteration can be used to converge to better depth estimates. Their plans for further work with this technique center around automating the point selection and initial estimating of their depths.

4.3.8 – Nagel and Neumann

"On 3D Reconstruction from Two Perspective Views," Hans-Helmut Nagel and Bernd Neumann, Seventh Int. Joint Conf. on Artificial Intelligence, Vancouver, B.C., p. 661-663, August 1981.

A registration scheme involving non-linear equations in the rotation parameters of the stereo camera transform is described. The registration scheme is derived using a perspective projection model, and the analysis depends upon there being given a set of image point matches. No discussion of the numerical procedures used in solving the non-linear system is presented. The derivation involves the same basic co-planarity condition discussed in the introduction and used also by [Clarkson 81]. Here the condition that point match cross-product vectors are co-planar (see discussion for [Clarkson 81]) is used to derive the mentioned equations: since all cross-product vectors should be co-planar, the cross-product of two of *these* vectors should be perpendicular to the other vectors, i.e., should have a dot product with them equalling zero. By such considerations, a second-order non-linear system in the rotation parameters is derived, which is adequate to determine the rotation. Five image point matches from two views of a scene are required.

4.3.9 – Neumann

"Exploiting Image Formation Knowledge for Motion Analysis," Bernd Neumann, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 6, November 1980.

The analysis described in this paper utilizes a static camera viewing a moving scene, with motion restricted to rotation about an axis normal to the line of sight (like a turntable). The process sets out to determine the angle of rotation (assuming it is constant and that viewing slices occur at uniform spacings). It is shown that four views of a point are sufficient for this. The system seems quite sensitive to intensity noise (it gives poor results if noise is over 2%).

4.3.10 – Nevatia

"Depth Measurement by Motion Stereo," Ramakant Nevatia, Computer Graphics and Image Processing, 5, 1976.

The author cites the tradeoff between wide baseline with higher positioning accuracy and small baseline with easier correspondence determination. The intent of the work described here is to improve stereopsis accuracy by tracking 'interesting' features (area based) through a series of views, and using extremal views for depth determination. It uses Normalized Mean Square, rather than Normalized Cross-correlation, saying that for the imagery used, the sensitivity to scene illumination didn't matter (difference techniques don't correct for gain/bias, whereas product NCC does). Search is limited to be along a line.

4.3.11 – Prasdny

"Relative Depth and Local Surface Orientation from Image Motions," K. Prazdny, Proceedings of the ARPA Image Understanding Workshop, p. 47-60, April 1981.

This paper is concerned not so much with finding the stereo camera transform as with avoiding the need for determining the transform altogether. It aims for solving directly for the relative depth (ratio of distances to camera center) of image points based on optic flow information. The reasoning involved is roughly as follows. We can consider the motion of a rigid object relative to a camera as having two components: a rotation about the camera center, and a translation. The motion of an object point can be considered at an instant to be that of a ray in space, instantaneously rotating about an axis through the center of projection, i.e., moving in a circle with an axis of rotation passing through the center of the circle and the camera center. This circular movement has some angular velocity, which has a rotational and a translational component. It turns out that all points on a rigid object have the same rotational component in their instantaneous angular velocities, so that the rotational component is independent of relative depth, and the difference of the angular velocities of two points on a rigid object is dependent only upon their respective translational components, which do depend on relative depth: these components are just inversely proportional to the object distance to camera center. Thus relative depth can be recovered as a ratio, if instantaneous angular velocities can be determined. Prazdny shows that the angular velocity of a point can be determined from its optic flow vector and the time derivative of that vector: the instantanous circle of motion of a point can be determined from the tangent to that circle recoverable from the optic flow vector, and the change of that tangent over time. Prazdny also discusses the determination of local surface orientation of objects with similar data, assuming the image FOE is known. Implementation issues are not greatly discussed, although it is pointed out that the optic flow must be known to a high degree of acccuracy for this method to succeed. On the other hand, there is a great deal of potential redundancy present, since relative depth values can be found for any pair of object points.

"A simple method for recovering relative depth map in the case of a translating sensor," K. Prazdny, Seventh Int. Joint Conf. on Artificial Intelligence, Vancouver, B.C., p. 698- , August 1981.

A method is given for producing a relative depth map of a scene given the optic flow, under the conditions that the FOE is known and the camera is translating (and not rotating) in a stationary world. Requires large local brightness changes, e.g., at edges. The method is a special case of that described in the paper just above.

"Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinear Moving Observer," K. Prazdny, Computer Graphics and Image Processing, 17, P. 238-248, 1981.

The method described uses optic flow information in perspective projection model to determine relative camera motion. The method involves the minimization of an error function of the rotational component of the optic flow, using the FOE consistency condition described in the introduction. In this case, the optic flow vector due to camera motion is considered to result from three vector components adding together, two of which are due to camera rotation. These components are derived from camera rotation parameters through geometrical considerations, so that an estimate of rotation parameters results in an estimate of optic flow components, which can be subtracted from observed optic flow to yield an estimate of the translational optic flow component. This component would result from a camera that is only translating, so all such components should point to the FOE. The extent to which they do not do so can be made the basis of an error function, though this is somewhat problematic when the FOE is not known. Here an error function is determined by choosing an arbitrary image point and determining the intersection of other optic flow rays with its optic flow ray. The dispersion of these points of intersection, as measured by the variance, is then a measure of the error of the estimate of the rotation component. Experiments with noiseless artificial data are mentioned, in which the minimization converged over a broad range of initial estimates.

4.3.12 – Roach and Aggarwal

"Determining the Movement of Objects from a Sequence of Images," John W. Roach and J. K. Aggarwal, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 6, November 1980.

This paper describes a solution to the problem of tracking the three-dimensional motion of an object. The solution depends upon solving a system of non-linear equations. It uses 2 views of six points or 3 views of four points with a modified least-squares error method. Overdetermination is critical to obtaining accurate results. Input is assumed to be a set of points representing a rigid bodied object. Points selected are given in correspondence across images - It does not address the problem of selecting or matching such points. The viewing camera is assumed to be stationary, watching smooth and continuous motions in the scene.

4.3.13 – Rashid

"Towards a System for the Interpretation of Moving Light Displays," Richard F. Rashid, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 6, November 1980.

This paper describes a system for the tracking and clustering of light points arising from jointed motions of a body. It uses a minimal spanning tree to cluster points for the grouping in 4-d space (x,y,Vx,Vy) of position and velocity in the image. A memory, or history, of previous attachments is included by adding to each link a function of the cost of the corresponding link in the previous frame. Cuts are selected wherever the cost of a link is over 50%greater than the cost of the two links nearest it ... this to separate objects. For intraobject structure, it subtracts the velocity of a cluster's centroid from those of its components, and allows equal but opposite velocities to be considered coupled (complementary motions). It is not clear how the integration of information over time is effected ... nor what the 'function' is that is applied to the previous frame's costs to affect the present frame values (how was this derived/determined?). The only aspect of interest here is the author's use of minimal spanning trees, but there is nothing substantially original in this technique or its use here.

4.3.14 – Thompson

"Combining Motion and Contrast for Segmentation," William B. Thompson, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 6, November 1980.

The research work reported here deals with using both motion and brightness information in segmenting images into regions. Assumed is a stationary camera and a changing scene, and two images are used in doing the segmentation. The technique is able to inform of translational motion of rigid objects (no rotations, scale changes, or deformations). Velocity estimates, obtained from the intensity gradient field of the image pair, and intensity similarity are used to make initial region clusters. Merges are made on the basis of both parameters. Many ad hoc thresholds and heuristics are used in the clustering/merging.

4.3.15 – Tsuji

"Tracking and Segmentation of Moving Objects in Dynamic Line Images," Saburo Tsuji, Michiharu Osada, Masahiko Yachida, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 6, Nov 1980.

To be used in cine-film understanding of cartoon figures, this process operates by segmenting the image into regions which are then matched over time (frame sequences). Structural descriptions of each object include properties, spatial relations, and motion patterns. A flexible template matching scheme is used to relate object parts across frames. Breaks in 2-D structure of the image line depictions tends to break the tracking. Multi-frame information is not used (it could provide a useful context for next pair analysis). Motions analyzed are strictly planar.

4.3.16 – Sobel

"Camera models and machine perception," Irwin Sobel, Stanford University Report, AIM-121, 1970.

The problem of determining certain numerical camera parameters is put in a mathematical setting.

4.3.17 – Ullman

"The Interpretation of Visual Motion," Shimon Ullman, MIT-AI Thesis, May 1977.

The sections of this thesis concerning registration (here in the context of determining the parameters of motion of a rigid object) describe results concerning the unique interpretation of input image point matches. Specifically, it is shown that, in a parallel projection model, 3 views of an object with 5 point matches suffice to determine the structure and the motion of the object. An algorithm is also given for recovering the motion given this kind of input. Numerical implementation issues are not addressed. Procedures are also described for the recovery of motion and structure parameters in a perspective projection model, under the restrictions of either pure translation, or rotation about a known axis combined with a translation.

4.3.18 – Webb and Aggarwal

"Structure from motion of rigid and jointed objects," J. Webb, and J. K. Aggarwal, Seventh Int. Joint Conf. on Artificial Intelligence, Vancouver, B.C., p. 686-691, August 1981.

A method for recovering a scene depth map given optic flows is presented. Assumes parallel projection of rigid jointed obiects, which are translating and rotating about a fixed axis. These assumptions allow the interpretation of motion with only two input points. Psychological data and implications are considered.

4.3.19 – Williams

"Depth from Camera Motion in a Real World Scene," Thomas D. Williams, IEEE Transactions on Pattern Analysis and Machine Intelligenc, Vol. 2, No. 6, November 1980.

This is a system to refine depth and orientation estimates of surfaces in a scene. The assumption is that all surfaces are either horizontal or vertical. Motion is used to determine distance. The focus of expansion (FOE) is known (although the thesis upon which this is based describes a technique for determining the FOE from the imagery). The scene is first segmented into surfaces ... i.e. regions so called (in one of the two examples regions were clustered interactively to form surfaces). It would seem to require *a priori* knowledge of the camera's relative motion, since the error function is the actual scene intensities minus a 'synthetic' estimation of the image, derived from the previous image and the 'surface model'.. it is not explained how this surface model is obtained (it would suggest that the scene be known before it can be analyzed!). Several 'synthetic' images are generated, each coming from a different assumption of surface heights and distances. So, the surfaces pump up and down, in and out, and the combination giving least error is selected as the best.

4.4 References

[Clarkson 81] Clarkson, K.L., "A Procedure for Camera Calibration," *Proceedings DARPA Image Understanding Workshop*, 175–177, April 1981. (cited on p. 164,170)

[Lucas 81] Lucas, Bruce D., and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Seventh Int. Joint Conf. on Artificial Intelligence*, Vancouver, British Columbia, 674–679, August 1981. (cited on p. 164)

[Nagel 81] Nagel, Hans-Helmut, and Bernd Neumann, "On 3D Reconstruction from Two Perspective Views," *Seventh Int. Joint Conf. on Artificial Intelligence*, Vancouver, B.C., 661–663, August 1981. (cited on p. 164)

[Prazdny 81a] Prazdny, K., "Relative Depth and Local Surface Orientation from Image Motions," *Proceedings of the ARPA Image Understanding Workshop*, 47–60, May 1981. (cited on p. 164)

[Wong 75] Wong, K.W., "Mathematical Formulation and digital analysis in close-range photometry," *Photogrammetric Engineering and Remote Sensing*, vol. 41, 1355, 1975. (cited on p. 165)

Chapter 5

CORRESPONDENCE CONSTRAINTS FOR STEREOPSIS

5.1 Introduction

One of the primary constraints used in stereo analysis is one afforded by knowledge of the imaging configuration. It is referred to in the photogrammetry literature as 'epipolar geometry'.

When the imaging relative orientations are known, the search for correspondences between left and right images can be restricted to a linear path. To determine the corresponding linear paths in a pair of images, the viewing geometry must be analyzed. First, consider a restricted camera geometry, referred to as the *'collinear'* camera model: here, the origin is located at the focus of the left camera and the right camera focus lies on the x axis. The two image planes are coplanar and are perpendicular to the z axis. The baseline, B, is the distance between focii. Given any point on an object, we define an "epipolar plane" as that plane passing through the object point and both focii. This plane intersects the two image planes, defining an "epipolar line" in each. These lines are parallel to the x axis in our *collinear* camera model, and we will refer to them as "epipolars". Corresponding points must lie on corresponding epipolars, that is lines with the same y-coordinate in both left and right images.

If a stereo image pair were taken with a different (non-collinear) geometry, a similar mapping could be obtained for corresponding (conjugate) epipolars, although they would be neither parallel to each other nor parallel to the scanning axis (x axis) of the cameras. [Hallert 1960, page 27] discusses general epipolar geometry.

Further constraints on the correspondence process tend to be either photometric or geometric in nature, as the following summaries will indicate. Marr and Poggio [Marr 1977] talk about ***uniqueness*** and ***smoothness*** in limiting correspondence search, Arnold and Binford [Arnold 1980] present geometrically-based estimation criteria for selecting corresponding image feaures (***edges)),*** [Ryan 1979] discusses the inherent inadequacies of photometric cross-correlation matching, and Mayhew and Frisby [Mayhew 1981] improve upon Marr and Poggio's specification of a stereopsis process with the introduction of figural continuity constraints and the use of cross-channel correspondences based upon 'primal sketch' type descriptive primitives ([Marr 1976a]).

5.2 Correspondence Constraint Summaries

5.2.1 – Marr and Poggio 1977

"A Theory of Human Stereo Vision," Marr, D. and T. Poggio, MIT Artificial Intelligence Memo No. 451, November 1977.

Marr and Poggio outline a theory for stereopsis. A simplistic characterization of their approach, one which is shared with all models of stereo vision, is to select a particular location on a surface in the image of one eye, identify it in the other image, and measure the disparity implied by the correspondence. Their work centers on the difficult first two steps here: defining the selection of pertinent locations in the images, and specifying the technique of the matching. They define pertinent locations in an image to be zero-crossings in a difference-of-gaussian convolution on the imago, and they use psychophysical evidence to justify the use of these spatial frequency tuned convolutions. Filters chosen correspond loosely with those felt to be present in human vision, although their frequency ranges are slightly larger. These zero-crossings from the various frequency ranges are matched across images. Statistical estimates on the characteristics of the filters employed are used to constrain the disparity limits of correlable zero-crossings within a frequency range. Two geometric observations further limit the allowable matchings. An assumption on the ***uniqueness*** of points in space is used make the mapping at most one-to-one – a point on a surface in space is seen at a particular location in one image, and it can have at most one appearance in the other image. An assumption on the ***continuity*** or ***smoothness*** of three-space provides the second matching constraint: matter is generally cohesive, with breaks between objects or surfaces statistically fairly rare, and this suggests that disparities should vary smoothly almost everywhere over a scene. Abrupt variations in depth along surfaces felt otherwise to be smooth are likely incorrect. The details of their theory were not specified precisely in this paper; a reading of the summary of Grimson's work ([Grimson 80]) will clarify many of the implementation issues.

5.2.2 – Arnold and Binford 1980

"Geometric Constraints in Stereo Vision," R. D. Arnold and T. O. Binford, Soc. Photo-Optical Instr. Engineers, Vol. 238, Image Processing for Missile Guidance (1980), p. 281-292

Arnold and Binford describe the derivation of two important analytic functions based on geometric constraints for the matching of edges along conjugate epipolar lines. One concerns a constraint on the ***intervals*** between adjacent edges and the other concerns a constraint on the ***orientation*** of matched edges. These results allow a distribution function in the three-dimensional object space to be translated to a distribution function in the two-dimensional image space. The image space functions allow probability estimates to be made on the likelihood that edges from the two images correspond. These estimates can be used in the selection of the best pairing of edges among a set of alternatives.

The analysis deals with the case of edge features on corresponding epipolar lines in the two images. The feature parameters of interest are the position and orientation of edges, that is, the points at which image edges intersect the epipolar, and the orientation of the edges at those points. The authors invoke the general assumption that edge and surface orientations are not related to observer position.

Their edge orientation analysis is based on a mapping of projective edge orientations in the image planes to a sphere of uniformly distributed surface normals in three space. The mapping of surface normals to image edge orientations at various imaging baselines defines the distribution function.

Their results lead them to conclude that in biological vision: a) stereo cells should show a half width of about 9 degrees for angle differences in the two eyes; and b) such stereo cells will be insensitive to angles of vectors in space. However, those angles can be calculated accurately at a later stage from associated vectors. This observation has not been seen in the literature, but they say experiment supports their interpretation. Nelson, Kato, and Bishop [Nelson 77] show stereo cells with orientation half widths from 10 degrees to 20 degrees, which are insensitive to space angles of vectors.

A similar mapping of surface orientations in three-space to their projected intervals in two-space leads to similar constraints on the distribution of interval correspondences.

Their analysis suggests that the two functions are sharply peaked even for the 60 degree vergence angles used in aerial photography. When angles corresponding to human vision are used, the conditions are extremely strong. This leads them to recommend the use of mapping sequences with small angle stereo. For example, instead of a pair of images with a 60 degree baseline, a sequence of 10 images at 6 degrees would provide the same overall baseline. Tight constraints would simplify the task for the program, which could track features from frame to frame, then make depth estimates based on the accumulated baseline (see [Nevatia 76]).

It is interesting to note that the underlying distribution assumptions that lead to the results they cite may be modified by known distributions, where they are available. To first order, the assumption of uniform distributions in the object space is useful. However, the functions can be made to incorporate knowledge of the scene when it is available. Cultural scenes, for obvious structural reasons, tend to be strongly oriented with respect to gravity. Here, horizontal and vertical surfaces predominate, and one would expect a strong bimodal surface distribution, and a similar bias in the distribution of their projective interval correspondences.

5.2.3 – Ryan, Gray and Hunt 1979

"Prediction of Correlation Errors in Stereo-Pair Images," T. W. Ryan, R. T. Gray, and B. R. Hunt, SIE/DIAL-79-002.

The authors objective is to investigate sources of digital correlation error and to develop image quality measures which can be used to predict the magnitude of correlation errors in a particular region of an image. This is with the long term aim of providing pre-screening capability that would allow defective imagery to be identified, and enabling manual processing or some sort of image enhancement for them. The two primary sources of difference between images are presented as resulting from *film or sensor noise* (a weakly signal-dependent random quantity injected into both images), and *relief distortion* (the effect of viewing orientation and perspective on the relative attitude of scene surfaces). A third source of image difference is mentioned as being that arising from the sensing averaging process, where discretization of the original scene intensity map over a fixed window leads to conjugate image points having differing intensities. Simulations of the image formation and correlation processes (with two different correlation measures.. the Cramer-Rao, and MSE) suggest that automated correlation may be slightly improved by low-pass filtering the images, and they reference a paper by Heleva which says that their exists a relatively narrow band of spatial frequencies at which such correlators will function best. Equally, smooth images have high error attributable to correlator 'self noise', particularly where the images have high signal-to-noise ratios and smooth correlation functions. The determination of correlation failure is dependent upon the relief distortion estimated for the scene. A feature-based simulation was also carried out, using as features over the window: the Cramer-Rao measure, brightness variance, contrast difference, contrast ratio, and brightness median absolute difference (MAD). The classification and recognition operations on the data were performed by commercially available software. Individual analyses of the various features indicated that MAD is generally best, with contrast ratio worst. Increasing window size improves feature performance. A final simulation feature test was the calculation of a multiple linear regression fit to the date (with an 8 by 10 averaging window and 11 by 3 feature window). The residual sum square error was then calculated for the various feature combinations various feature combinations. The results indicate that variance with the normalized Cramer-Rao measure provides the best fit. The Cramer-Rao measure is a lower bound on the variance of any unbiased estimate of the parallax in an image-pair region, although its use presupposes an estimate of the noise spectral density (which can be obtained by analyzing a structure-free part of the image). The experiments here suggest that simple contrast measures are just as good at estimating correlation errors as is the Cramer-Rao bound, although the latter is theoretically more sensitive.

5.1.3 – Mayhew and Frisby 1981

"Computational and Psychophysical Studies Towards a Theory of Human Stereopsis," John E. W. Mayhew and John P. Frisby, Artificial Intelligence, 1981.

This paper discusses a significant variant to the theory of stereopsis posed by Marr and Poggio [Marr 1977]. The authors show that human vision utilizes more than simply zero-crossings in obtaining local stereopsis; rather, they suggest that peaks in the signal (difference of gaussians) also be used in the correspondence, and cite compelling psychophysical evidence in support of this. They are saying that luminance variation, and not just luminance discontinuity, plays a role in stereopsis. The authors feel that the goal of stereo processing is the construction of a 'Binocular Raw Primal Sketch (BRPS)', and feel this is attained through the interaction of local matching and global disambiguation. The global disambiguation is effected through the use of *figural continuity* and *cross channel correspondences.* Figural continuity capitalizes on the connectedness of zero-crossings in the two projective images, and with the restriction that connected edges have connected correspondencees, implicitly maintains the Marr and Poggio constraint that zero-crossings be smooth in three-space. Their definition of "connectedness", however, is not precise. A distinction they draw on the lack of 'explicit' reference to disparity information seems vacuous, in that disparity limits are implicit in their figural continuity, and through an explicit utilization, can be evaluated quantitatively. The correspondence algorithm they mention allows a feature in one of the images to match *several* features in the other image. This is in sharp contrast to the matching algorithm of Marr and Poggio, where edge correspondencees are at most one-to-one. This enables stereopsis of Panum's limiting case, but brings an untold increase in complexity to the matching process. Again, they limit correspondences to zero-crossings of same sign, and of roughly similar orientation, and this without adequate statistical analysis.

5.2.5 – Liebes 1981

"Geometric Constraints for Interpreting Images of Common Structural Elements: Orthogonal Trihedral Vertices," S. Liebes Jr., Proceedings of the ARPA Image Understanding Workshop, 1981.

This research deals with the use of knowledge in a specific domain to guide the stereo correspondence process. Orthogonal trihedral vertices (OTV's) are ubiquitous in cultural scenes. Constraints within the domain of orthogonal trihedral vertices can be applied to the stereo analysis of cultural scenes: the faces and edges of OTV's of single objects and often of collections of objects tend to be mutually aligned as well as gravity aligned; edge vanishing points my be inferred from perspective views; the object space invariance of vanishing points constrains the projections in stereo image pairs. This paper proposes a mechanism for facilitating correspondence matching and orientation determination for planar surfaces on OTV's.

5.3 References

[Arnold 80] Arnold, R.D., and T.O. Binford, "Geometric Constraints in Stereo Vision," *Soc. Photo-Optical Instr. Engineers*, vol. 238, Image Processing for Missile Guidance, 281–292, 1980. (cited on p. 176)

[Grimson 80] Grimson, W.E.L., "Computing Shape Using a Theory of Human Stereo Vision," Department of Mathematics, MIT, June 1980. (cited on p. 177)

[Hallert 60] Hallert, Bertil, *"Photogrammetry, Basic Principles and General Survey,"* McGraw-Hill Book Company Inc., 1960. (cited on p. 176)

[Marr 76a] Marr, D., "Early Processing of Visual Information," *Philosophical Transactions of the Royal Society*, London,Series B, 275, p 483–524, 1976. (cited on p. 176)

[Marr 77] Marr, D. and T. Poggio, "A Theory of Human Stereo Vision," MIT Artificial Intelligence Memo no. 451, November 1977. (cited on p. 176,179)

[Mayhew 81] Mayhew, John E.W. and John P. Frisby, "Computational and Psychological Studies Towards a Theory of Human Stereopsis," *Artificial Intelligence Journal*, vol. 16, 1981. (cited on p. 176)

[Nelson 77] Nelson, J.I., H. Kato and P.O. Bishop, "Discrimination of orientation and position disparities by binocularly activated neurons in cat striate cortex," *Journal of Neurophysiology*, 40(2):260–283, March 1977. (cited on p. 178)

[Nevatia 76] Nevatia, Ramakant, "Depth Measurement by Motion Stereo," *Computer Graphics and Image Processing*, 5, 1976. (cited on p. 178)

[Ryan 79] Ryan, T.W., R.T. Gray, and B.R. Hunt, "Prediction of Correlation Errors in Stereo-Pair Images," SIE/DIAL–79–002. (cited on p. 176)

Chapter 6

HARDWARE AND ARCHITECTURES FOR COMPUTER VISION

6.1 Overview

6.1.1 – Separation of tasks

To discuss hardware and architectures for computer vision, it is convenient to divide the necessary tasks into low-level vision and symbolic reasoning. Although the division is somewhat artificial, the intuitive notion is that low-level vision encompasses the same tasks as the first few stages of the (human) visual system, and symbolic reasoning is concerned with using the results of low-level vision operations to extract spatial information from images. More detail can be found in the other sections of this survey, but the most relevant point to this section is that low-level vision processes are characterized by their regular, relatively simple nature. Low-level vision processes have an inherent high degree of parallelism due to their local nature (each process is repeated for all pixels in the image, and the computation at each pixel only involves data in a small neighborhood around that pixel). The paralellism in symbolic reasoning is better expressed in terms of the algorithms involved rather than the image — for example, symbolic reasoning involves some searching, which can be done along each branch in the search tree in parallel.

6.1.2 – General-purpose computing

One of the ways to speed up both low-level vision and symbolic reasoning is to speed up the operation of general-purpose computers. This is usually done using pipelining and special high-bandwidth memory designs [Levine 82]. Another approach that is gaining popularity is to make machines in which special-purpose operations can be microcoded with relative ease ([S-1 79], [Thacker 79], [Lampson 81]). The main difficulty with building faster general-purpose computers is that the resulting "supercomputers" usually only run about ten times as fast as widely available computers.

Some high-speed general-purpose computers attempt to take advantage of any parallelism that is inherent in the algorithms they execute (in other words, the computer discovers the parallelism while it is running rather than specifically being programmed for parallel operation. [S-1 79] describes a multiprocessor machine whose operating system and compilers coöperate to schedule parallel tasks in 16 uniprocessors, each of which is highly pipelined to take advantage of the parallelism found in a single instruction stream. Another scheme for taking advantage of parallelism in general-purpose algorithms is the use of data flow architectures ([Ackerman 82], [Davis 82]). Data flow architectures execute programs that are written in terms of the relationships between computed data rather than in terms of the sequence of operations needed to compute the data. Since data flow programs explicitly show all data dependencies in the computation, they can be executed at the maximum rate (for a suitable definition of "maximum"), given enougn processors. However, the main limitations of data flow architeçtures [Gajski 82] are the fact that can only execute purely applicative languages (languages with no side effects), the difficulties encountered in implementing them efficiently, the difficulties of making them deal effectively with arrays, and the fact that they seem to only take advantage of very fine-grained parallelism (the same sort of parallelism exploited by pipelining).

6.1.3 – Symbolic reasoning

Since much of the software in existence for doing symbolic reasoning is written in LISP, a reasonable approach to doing faster symbolic reasoning would be to make a machine to run LISP very fast. The Lisp Machine [Waller 81] takes this general approach, but is more like a personal computer than a high-performance vision machine. The Japanese have proposed a high-performance architecture for performing LISP-like tasks — their so-called "fifth-generation" computers ([Manuel 81a], [Manuel 81b]). The problem with these approaches is due to the fact that making a fast LISP machine is much like making a fast general-purpose computer, so once again, advanced architectures will only be about ten times as fast as simpler architectures with the same circuit technology.

6.1.4 – Low-level vision

Low-level vision, on the other hand, lends itself to tremendous speedup due to its inherent parallelism. Several different architectures can be used to take advantage of this parallelism: vector processors, highly pipelined special-purpose processors, systolic processors, multiprocessor configurations, and processor arrays.

Vector processors

Vector processors are designed to perform operations efficiently on vectors by using special memory configurations, multiple arithmetic units, and highly-pipelined data paths. These machines reach peak performance when the first several elements of a vector have been processed and the pipelines are filled with data. Examples of high-performance vector machines are the Cray-1, Cray-2 [Iversen 81], and the Cyber 205. A relatively recent phenomenon has been the introduction of low-cost vector processors that are used as peripherals to minicomputers. These processors (called array processors) have received much attention due to their high performance/price ratio ([Charlesworth 81], [Maron 81]).

Highly pipelined special-purpose processors

Highly pipelined special-purpose processors are usually built using conventional design techniques to do one specific task. For example, [Eversole 80] describes a special-purpose pipelined processor for doing dot products. These architectures are usually impressively fast doing the task for which they are designed, but are quite inflexible.

Systolic processors

So-called "systolic" arrays are an extreme case of highly pipelined special-purpose processors [Kung 80]. The arrays are usually made from units that calculate $c' = c + ab$ and are connected in such a way that the processor interconnections directly model the data flow of the problem. Since one processor is used for each computation in the problem, a tremendous amount of pipelining is achieved. In addition, the regular structure and interconnection of the processors makes these arrays very suitable for VLSI implementation. Unfortunately, systolic arrays are inflexible in both the algorithms they execute and the size of the input data, and not all algorithms are suitable for implementation in systolic arrays (although [Guibas] shows some non-numerical applications).

Multiprocessor configurations

Due to the increasing power of microprocessors, several "multi-microprocessors" have been constructed — these are networks of low-power computers, providing relatively large computing power (for parallel tasks) at a fairly low price. One of the first of these was C.mmp [Fuller 76], a shared-memory network of PDP-11 minicomputers. ZMOB, a ring network of Z-80 microcomputers, has recently been built and has had some image processing algorithms implemented on it [Kushner 80].

Processor arrays

Processor arrays seem to be the most promising architectures for low-level vision. Processor arrays are conceptually similar to multiprocessors, but usually have a larger number of simpler processors. More importantly, a processor array is controlled by a central controller that broadcasts the same instruction to all processors. In addition, each processor usually has a local memory, and communication between processors is done through a nearest-neighbor network rather than a shared memory. One of the earliest processor arrays was the Illiac IV ([Barnes 68], [Kuck 68]). Although the Illiac was used primarily for aerodynamic simulation [Chapman 75], it would perform fairly well on low-level vision tasks (although the full power of its floating-point capabilities might not be exercised). The trend in recent years has been to make larger arrays of simpler processors. The CLIP processor ([Duff 77], [Duff 78]) uses extremely simple processors optimized for single-bit calculations, and in addition has a feature allowing it to propagate single-bit computations through the array asynchronously. The SAM array [Blank 81] is similar to the CLIP although it is a more conventional synchronous design and is optimized for a limited number of design-automation algorithms. The DAP processor [Marks 80] has a more complicated processor designed to do bit-serial arithmetic on multiple-bit data using more conventional synchronous design. The MPP ([Batcher 79], [Batcher 80]) is also a bit-serial machine with variable word size, but has much more memory per processor, and is correspondingly much larger physically. [Lowry 81] describes a bit-serial processor array optimized for 8-bit data in which the processor speed is independent of the array size, and discusses possibilities of implementation of the array using wafer-scale integration.

6.1.5 – Software

Of course, parallel architectures will do nothing without software to run them. [Kung 80] gives a good overview of the kinds of software that can be designed for parallel architectures, and

[Kuck 76] describes methods for converting serially-described algorithms into parallel ones. [Brode 81] and [Karplus 81] discuss some of the software issues encountered when using array processors. The best description of software issues as far as computer vision is concerned is [Marks 80], since it describes vision algorithms that have been implemented on working hardware.

6.2 Hardware and Architecture Summaries

6.2.1 – Barnes et al. - Illiac IV

"The ILLIAC IV Computer," Barnes, George H., Richard M. Brown, Maso Kato, David J. Kuck, Daniel L. Slotnick, and Richard A. Stokes, IEEE Transactions on Computers, Vol. C-17, No. 8, August 1968.

Summary

This paper describes the ILLIAC processor array that was designed as an extension of the SOLOMON processor array. The ILLIAC is four arrays of processors; each array having a central controller and 64 processors arranged in an 8 × 8 configuration. Each processor has 2K 64-bit words, four 64-bit data registers, a 64-bit arithmetic unit, a 16-bit index register and adder, and an 8-bit mode register. The central controller executes instructions from processor memory through a cache. These instructions can either deal with operation of the controller itself or cause processor action. A processor's arithmetic unit can do one 64-bit floating-point operation, two 32-bit floating-point operations, eight 8-bit fixed-point operations, one 64-bit boolean operation, and either a 64-bit or 32-bit barrel shift. Operands for the arithmetic come from the data registers, which can be loaded either from the processor memory (indexed by the index register) or the R-register of an adjacent processor (in a 4-neighbor configuration). The mode register can be set by data-dependent operations and can be used to disable processor operations, allowing conditional operation of sections of the array.

Commentary

Since this article is so old, it is easy to criticize it from a historical perspective, especially since only one fourth of the machine was built and even then it was several years behind schedule. However, such criticisms should be tempered with the realization that ILLIAC is still the fastest computer in existence for certain computations and it was used heavily until its recent (late 1981) demise for a large amount of aerodynamic simulation. In addition, it had one of the highest-bandwidth disk memories ever built (necessary to keep the processors supplied with data at a reasonable rate) and it certainly had one of the largest memories (128K 64-bit words) of any computer in the early 1970's.

This particular article suffers from a lack of concrete examples to show how the processors can be used to carry out parallel computations. In addition, it is unclear how some of the operations are controlled and what the exact instruction format is. For example, it is never clear how a program can cause a processor to access data from a neighbor.

The whole idea that parallel algorithms can be fit into a 4-neighbor arrangement is questionable, although certainly many image-processing and computer vision algorithms can be put into parallel form this way. The article even mentions that the ILLIAC's four independent arrays can run independently or be coupled as two groups of two or one group of four in order to better match the size of the problem.

6.2.2 – Brode

"Precompilation of Fortran Programs to Facilitate Array Processing," Brode, Brian, Computer, Vol. 14, No. 9, September 1981.

Summary and Commentary

The author describes a compiler called "VAST" (which is apparently a successor to one called "FAST") that takes Fortran programs and optimizes them for running on "any" vector machine (it currently only generates code for one machine, which isn't specified).

Several different types of vector optimization are discussed, but all of them are somewhat obvious. Right at the end of the article, some more non-trivial optimization is mentioned, but disappointingly enough, it is only mentioned in passing. This article would be a reasonable quickie introduction but has no solid content.

6.2.3 – Charlesworth

"An Approach to Scientific Array Processing: the Architectural Design of the AP-120B/FPS-164 Family," Charlesworth, Alan E., Computer, Vol. 14, No. 9, September 1981.

Summary

The author describes the architecture of a peripheral array processor, contrasting it with a vector processor. This particular processor was designed using inner-product and row-elimination as "kernel benchmarks", so it has one adder and one multiplier (it is mentioned that empirical studies of FORTRAN programs would suggest having two adders and one multiplier). The pipeline length is reduced to three stages to achieve a balance of 3:1 in execution speed of vector:scalar operations (due to a study by Amdahl showing that computations can be about 75%vectorized). The processor has a large main memory, a faster but smaller auxiliary memory, and two very fast registers. Everything is connected together in a switching arrangement which is not quite a full crossbar switch.

Three main programming methods are presented:

Sequential programming, in which microcode is used to "simulate" conventional machine instructions like move from memory to register, add register to register etc.

Overlapped programming, which is like sequential programming except that non-interfering instructions are overlapped.

Pipelined programming, which "vectorizes" a computation in a loop by computing result i through result $i+n$ at the same time using n functional units each at a different stage of computation. This also involves code to initialize and finish the loop.

Commentary

It is interesting to note taat available compilers generate overlapped code but not pipelined code.

If the AP-FPS processors are as successful as the author would like you to think, it suggests that extensive study of vision algorithms would be the best way to tailor fast but general-purpose hardware to them.

6.2.4 – Duff and Watson - CLIP 3

"The Cellular Logic Array Image Processor", Duff, M. J. B., and D. M. Watson, The Computer Journal, Vol. 20, No. 1, 1977.

Summary

This article describes the CLIP 3 array processor. CLIP 3 is a 16 × 12 array of processors optimized for single-bit operations. These processors are connected in an 8-neighbor configuration. Each processor has a one-bit A register (used to accumulate results), a one-bit B register (used to inject propagation signals), and 16 one-bit D memory locations (used to store intermediate results). Each processor can sum the interconnection bits from enabled neighbors (yielding a number between 0 and 8 inclusive) and tareshold this sum to generate one bit called T. (Interconnection bits from the edges are connected to a single "E" bit.) The T signal is then OR'ed with the B register to generate P, the "propagation" signal. Any arbitrary boolean function of A and P can be programmed to generate both the bit stored into the D memory and the bit passed to neighbors. Note that there is a combinational logic path through the processor from the neighbor interconnections to the interconnection generation, so the array must be allowed to "settle" before storing results in the D memory. The contents of the A registers can be loaded using a light pen on an oscilloscope display, and the contents of both A and B can be displayed on the oscilloscope. A central controller broadcasts instructions to all the processors by executing from a 256-word 24-bit instruction stream (probably one instruction every 3 μsec, although this is not clear). The AND of all the data bits to be written into the D memories can be used for conditional branching by the central controller.

The paper mentions that a CLIP 4 array is planned using LSI technology to put 8 processors on one chip and build a 96 × 96 array.

The paper gives two examples of how the array can be programmed. The first example finds the outer edges of objects by starting a propagation bit from the edge in all directions (using the E bit with all neighbor interconnections enabled) and stopping the propagation when a "1" in the A register is encountered. The result is then simply the AND of the A register and the propagation bit P. The second example, which is slightly more complicated, determines whether or not all areas in the array are connected. Then four programs are mentioned without detailed algorithms being given: extraction and histogramming of parameters (area, perimiters, size of horizontal and vertical projections) of binary images, reduction of shapes to single-pixel-wide line skeletons, solution of Laplace's equation for electrostatic potential (with 5-bit numbers), and spatial frequency lowpass filtering by expanding and shrinking objects.

Commentary

The individual processors in the CLIP system seem to be pretty wimpy, having only 18 bits of memory and taking 3 μsec to compute one result. In addition, the "sum and threshold" processing seems dubious, and isn't used at all in any of the examples given in the paper (the threshold is always set at zero). The idea of allowing data to propagate asynchronously through the entire array is a novel one, but suffers from the fact that the length of time needed to compute a result is indeterminate — it is possible to have a computation whose critical path goes through every processor in the array, and worse still, it is possible to have a computation whose results never "settle" at all. Even ignoring these pathological cases, the computation time must get longer as the array grows in size.

The examples given in the paper are difficult to understand. The assembly-like language used is nearly unreadable, and the labeling of the processor interconnection scheme must be discovered by the reader. Worse still, there appears to be a bug in the only non-trivial program presented.

6.2.5 - Duff - CLIP 4

"Review of the CLIP Image Processing System," Duff, M. J. B., Proc. National Computer Conference, 1978.

Summary

This paper describes the CLIP 4 computer. The purpose of the CLIP 4 is to be a general-purpose computer optimized for a typical range of image-processing operations. The processor for CLIP 4 is similar to that of the one for CLIP 3 except that the sum-and-threshold logic on the processor interconnection is replaced with a logical OR, the D memory is expanded to 32 bits, and additional logic exists for carry generation and propagation. In addition, there is a special facility for counting the number of ones in the A registers that replaces the similar OR function in CLIP 3.

The processors can be used to do arithmetic in a bit-serial fashion (with each processor storing all bits of both numbers) or in a column-wide fashion (with each processor holding one bit of both numbers. In the bit-serial mode, the carry wrap-around is used as an input to the interconnection network, whereas in the column-wide mode the processors are interconnected for carry propagation. Each processor can also do single-bit boolean operations (with no carry). Propagating operations (which are done asynchronously, as in CLIP 3) can either be set up as "simple" operations (which do not involve the B register) or "labelled' operations (in which the B register is OR'ed with the propagating bit).

At the time the paper was written, an 8-processor n-MOS chip had been designed and a 96 × 96 prototype array was planned for late 1978 or early 1979.

Commentary

The assembly language used to program the CLIP 4 is certainly much easier to understand than that used for the CLIP 3. This article is easy to understand in that the examples are simple and the different modes of using the processor are discussed separately. However, there are several errors in the text, and no discussion is given about how different modes are implemented using the hardware of the processor. In addition, the text mentions that the instruction time must be lengthened by 1.2 μsec for each propagation step through a cell, but it does not mention how the controller figures out how long a complete propagation takes.

6.2.6 – Eversole and Mayer

"Investigation of VLSI Technologies for Image Processing," Eversole, William L., and Dale J. Mayer, Proc. Image Understanding Workshop, April 1980.

Summary

The authors present a way of doing dot products using what they call "distributed arithmetic". Basically, in order to find the sum of a_ix_i, they take all the MSBs of the x_i and use them to address a lookup table containing the various sums of a_i, take this result and shift it left one, add in the result from the next most significant bit, and keep shifting and adding. This finds the result of a B-bit I-number dot product in B steps using 2^I memory. They have some other schemes for reducing memory bandwidth.

They also mention that they are planning to make a special-purpose VLSI chip to do this computation, and that they have a breadboard model working.

Commentary

This approach seems OK for image processing applications where a single linear computation is repeated many times. The authors claim that their scheme can be generalized to nonlinear functions (i.e. finding the sum of arbitrary $f(x_i)$) but it turns out the way they do this is by scrapping the distributed arithmetic idea, using lookup tables to compute the $f(x_i)$ in parallel and then adding them together with an adder tree.

6.2.7 – Guibas and Thompson

Guibas, Leo, H. T. Kung, and Thompson, "Direct VLSI Implementation of Combinatorial Algorithms," extended abstract.

Summary and Commentary

Presents "systolic" way of doing dynamic programming with a triangular array, and transitive closure with a 3-pass algorithm through a rectangular array. Supposedly the processing elements are "product accumulators" just like all the other systolic stuff that Kung has done.

One clever thing is that for both of these algorithms all cells have to do something special at a time proportional to their distance from the edge of the array using some measure. This is done by propagating the control along with the data in a wave-like manner.

6.2.8 – Karplus and Cohen

"Architectural and Software Issues in the Design and Application of Peripheral Array Processors," Karplus, Walter J., and Danny Cohen, Computer, Vol. 14, No. 9, September 1981.

Summary and Commentary

This article discusses peripheral array processors: what they are, what they are used for, why they are fast, and some of their problems.

Although the text of the article uses only the AP-120B while discussing examples, there is a table that makes a comparison of some commercially available array processors (although the authors admit that the table is made directly from advertising literature).

The authors say that the biggest stumbling block for array processors is availability of high-level programming languages to use with them. They mention an "AP Fortran" that comes with the AP-120B, but say that it can end up taking about four times as much time as a hand-microcoded routine to do the same task.

6.2.9 – Lowry and Miller

"A General Purpose VLSI Chip for Computer Vision with Fault-tolerant Hardware," Lowry, Michael R., and Allan Miller, Proc. ARPA Image Understanding Workshop, April 1981.

Summary

The authors describe work in progress on a wafer-sized array processor made up of a large number of bit-serial arithmetic units, each unit responsible for *computations on one pixel*. Each unit contains two 8-bit data registers, a 16-bit data register, and a full adder with a carry feedback bit. The data registers are configured as shift registers that can feed into the adder. The inputs to the adder can be selected from any data register or the adder output of an adjacent processor (in a 4-neighbor network). The output of the adder can be stored in any data register. A processor enable bit enables data shifting and storage and can be loaded from the adder output. The result is that each processor is fairly general-purpose. Algorithms for convolution, cross-correlation, thresholding, and zero-crossing detection are given.

A discussion of the problems involved in making an image-sized array is given, along with solutions to the problems that are under investigation. The authors mention that a single-processor prototype is working. and estimate that a 128 × 128 array on a single wafer is feasible using current technology. A performance improvement over a conventional computer of 2500 is estimated for vision and image-processing algorithms.

Commentary

The main drawback in the design appears to be its lack of memory. Each processor only has 34 bits of memory, which somewhat limits the amount of useful computation it can do before moving data to and from another processor. Another limitation which plagues all rectangular arrays of processors (such as the Illiac) is the problem of what to do with algorithms that either have non-local data interaction or are too large to fit in the fixed array. The authors do, however, state that the device is intended for local operations on an image that fits in the available processors. Although the paper seems to address the problems of implementing a full array in adequate detail, the fact that no wafer-sized chips exist indicates that unforseen problems may exist.

6.2.10 – Marks

"Low-level Vision Using and Array Processor", Marks, Philip, Computer Graphics and Image Processing, Vol. 14, 1980.

This article describes the DAP (distributed array processor) and gives details about implementation of several computer vision algorithms on it.

The DAP is a 32 × 32 array of one-bit processors (with plans for a 64 × 64 array). Each processor has an ALU and a 1K by one bit memory (with plans for a 4K by one bit memory). The ALU has three one-bit registers: the Q register for accumulating results, the C register for holding the carry in a bit-serial addition, and the A or "activity" register for conditionally disabling stores into the memory. The ALU can do a full addition, half addition, or logical AND, and can complement some of the operands. Operands to the ALU can be selected from adjacent processors in a 4-neighbor arrangement. In addition, there are "row highways" and "column highways" which can either be used to broadcast constants to the processors or can be used in a wire-AND fashion to read from the processors. The rows of processors can also be used to do word arithmetic, since there is a "ripple-carry" connection directly between the adders of each row.

The array is controlled by an MCU (master control unit) that executes instructions from memory through a cache. The MCU selects registers, addresses the memory, and controls the ALU for all processors (every processor executes the same instruction). The MCU also has eight registers connected to the row and column highways.

The DAP memories are in the address space of the host computer (one word in the host is one bit of the memories of one row of processors). The MCU appears as a peripheral to the host.

The author works with 192 × 192 pixel images in this paper; in order to use the 32 × 32 DAP, each processor must handle more than one pixel. The image could be divided into 32 × 32 pixel sections, with each section being mapped onto the processor array, but in order to minimize gross data boundaries within the array, the algorithms are implemented by segmenting the image into a 32 × 32 grid of square subimages, with each subimage mapped into one processor.

Representative algorithms from several important classes have been programmed on the DAP: region finding, edge finding, line finding, and higher-level scene interpretation. The region finding algorithm uses the DAP to histogram gray levels in the image. The host computer then uses the histogram to compute thresholds for separating the regions. Presumably the DAP can be used to actually separate the regions, but this is not stated explicitly. The edge finding algorithm is O'Gorman's Walsh function technique. The algorithm uses a 4 × 4 window; this has an adverse interaction the the 6 × 6 pixel mapping used to map the 192 × 192 image onto the 32 × 32 DAP. For line finding, the author suggests using a relaxation method (as described by Zucker), but shows that the DAP processors do not have enough memory to effectively implement the algorithm. Instead, an algorithm is used that finds common boundaries in the regions found by the region finding algorithm and does a least-square type fit of points to these lines, discarding points whose distance from the line are above an arbitrary threshold. An edge-tracking algorithm is also described, but it fails to make much use of the parallelism of the DAP. At the end of the article, the author mentions that parts of Waltz' scene-labeling algorithm have been coded in a parallel fashion.

Commentary

The best thing about this paper is that it presents working algorithms that run on existing hardware. It is unfortunate that all of the algorithms are written in assembly language, but the author does mention that a "DAP Fortran" is in a developmental stage but is not yet flexible enough to deal with the sorts of optimizations needed for the algorithms described.

The DAP interface is more or less inextricably linked to the host computer, since the processor array's memory appears in the address space of the host. A better alternative might have been to connect the DAP as a peripheral, although the author mentions that the m-mory-mapped alternative was chosen to allow other peripheral devices to load processor memory directly. Another problem with the memory-mapped scheme of addressing the DAP is that it will seriously infringe on the host's address space if the DAP ever becomes much larger, and to directly implement the current scheme the width of the DAP (number of processors in one row) must be the same as the word size of the host.

The author does not use the "vector mode" of the DAP (in which the rows are connected together as a 32-bit adder with ripple carry) in any of the algorithms discussed in the paper. Using this mode ties the data format to the size of the array, which is probably not a desirable feature (since it means that expanding the array affects the software). In addition, expansion of the array requires slowing down instruction execution speeds in order to allow the carry to propagate through the longer row. Certainly, implementing large fixed-point adders with ripple carry is a dubious practice at best.

Although it doesn't affect the hardware aspects of the paper, the region finding scheme implemented seems rather sensitive to gain and contrast as well as smooth shading variations, since it depends on absolute intensities in the image. This also affects the robustness of the line finding algorithm since it uses the results of the region finding algorithm. In addition, the line finding algorithm cannot be easily extended to curved lines, and it depends on a rather *ad hoc* threshold for discarding points "not on the line".

6.2.11 – Maron and Brengle

"Integrating an Array Processor into a Scientific Computing System," Maron, Neil, and Thomas A. Brengle, Computer, Vol. 14, No. 9, September 1981.

Summary

In this paper, the authors describe some of the considerations they had to face when using an AP-190L.

One consideration was the length of time involved in transfering data to the processor. Rather than operating directly from the host computer's memory, the array processor has its own cache. This necessitates a DMA transfer from one memory to the other, with a format conversion along the way. In addition, the interaction of two different floating-point formats must be considered.

Another consideration was the timing of processor commands. It turned out to save a great deal of time by "chaining" processor commands, that is, buffering them and doing a group of them with one system call.

6.2.12 – Thacker et al., Alto

"Alto: A Personal Computer," Thacker, McCreight, Lampson, Sproull, Boggs, CSL-79-11, Xerox PARC, August 1979.

Summary

This document describes the Alto hardware/software environment.

This is an interesting design which utilizes a microprogrammed multi-tasking processor to simplify I/O control by sharing the processor among all peripherals and substituting microprogramming for I/O control hardware. This seems to work well because the bottleneck in most computer configurations is the memory bandwidth (distributing I/O control doesn't help because only one controller can be using the memory at one time anyway).

Another point of interest is the way the hardware and software were designed more or less at the same time, simplifying their interface.

6.3 References

[Ackerman 82] Ackerman, William B., "Data Flow Languages," *Computer,* Vol. 15, No. 2, February 1982. (cited on p. 182)

[Batcher 79] Batcher, K., "MPP-A Massively Parallel Processor," *Proc. International Conference on Parallel Processing,* August 1979. (cited on p. 184)

[Batcher 80] Batcher, Kenneth E., "Architecture of a Massively Parallel Processor," *Proc. 7th Annual Symposium on Computer Architecture,* May 1980. (cited on p. 184)

[Barnes 68] Barnes, George H., Richard M. Brown, Maso Kato, David J. Kuck, Daniel L. Slotnick, and Richard A. Stokes, "The ILLIAC IV Computer," *IEEE Transactions on Computers,* Vol. C-17, No. 8, August 1968. (cited on p. 184)

[Blank 81] Blank, Tom, Mark Stefik, and Williem vanCleemput, "A Parallel Bit Map Processor Architecture for DA Algorithms," *Proc. 18th Design Automation Conference,* 1981. (cited on p. 184)

[Brode 81] Brode, Brian, "Precompilation of Fortran Programs to Facilitate Array Processing," *Computer,* Vol. 14, No. 9, September 1981. (cited on p. 185)

[Chapman 75] Chapman, Dean R., Hans Mark, and Melvin W. Pirtle, "Computers vs. Wind Tunnels for Aerodynamic Flow Simulations," *Astronautics and Aeronautics,* April 1975. (cited on p. 184)

[Charlesworth 81] Charlesworth, Alan E., "An Approach to Scientific Array Processing: the Architectural Design of the AP-120B/FPS-164 Family," *Computer,* Vol. 14, No. 9, September 1981. (cited on p. 183)

[Davis 82] Davis, Alan L., and Robert M. Keller, "Data Flow Program Graphs," *Computer,* Vol. 15, No. 2, February 1982. (cited on p. 182)

[Duff 77] Duff, M. J. B., and D. M. Watson, "The Cellular Logic Array Image Processor", *The Computer Journal,* Vol. 20, No. 1, 1977. (cited on p. 184).

[Duff 78] Duff, M. J. B., "Review of the CLIP Image Processing System," *Proc. National Computer Conference,* 1978. (cited on p. 184)

[Eversole 80] Eversole, William L., and Dale J. Mayer, "Investigation of VLSI Technologies for Image Processing," *Proc. Image Understanding Workshop,* April 1980. (cited on p. 183)

[Fuller 76] Fuller, Samuel H., "Price/Performance Comparison of C.mmp and the PDP-10," *Proc. 3rd Annual Symposium on Computer Architecture,* January 1976. (cited on p. 184)

[Gajski 82] Gajski, D. D., D. A. Padua, D. J. Kuck, and R. H. Kuhn, "A Second Opinion on Data Flow Machines and Languages," *Computer,* Vol. 15, No. 2, February 1982. (cited on p. 183)

[Guibas] Guibas, Leo, H. T. Kung, and Thompson, "Direct VLSI Implementation of Combinatorial Algorithms," extended abstract. (cited on p. 184)

[Iversen 81] Iversen, Wesley R., "Total Immersion Cools Supercomputer Logic," ***Electronics,*** Vol. 54, No. 24, November 1981. (cited on p. 183)

[Karplus 81] Karplus, Walter J., and Danny Cohen, "Architectural and Software Issues in the Design and Application of Peripheral Array Processors," ***Computer,*** Vol. 14, No. 9, September 1981. (cited on p. 185)

[Kuck 68] Kuck, David J., "Illiac IV Software and Application Programming," ***IEEE Trans. on Computers,*** Vol. C-17, No. 8, August 1968. (cited on p. 184)

[Kuck 76] Kuck, David J., "Parallel Processing of Ordinary Programs," ***Advances in Computers,*** Vol. 15, 1976. (cited on p. 185)

[Kung 80] Kung, H. T., "The Structure of Parallel Algorithms," ***Advances in Computers,*** Vol. 19, 1980. (cited on p. 184)

[Kushner 80] Kushner, Todd, Angela Y. Wu, and Azriel Rosenfeld, "Image Processing on ZMOB," TR-987, University of Maryland Computer Science Center, December 1980. (cited on p. 184)

[Lampson 81] Lampson, Butler W., and Kenneth A. Pier, "A Processor for a High-Performance Personal Computer," appeared in "The Dorado: A High-Performance Personal Computer, Three Papers," CSL-81-1, Xerox PARC, January 1981. (cited on p. 182)

[Levine 82] Levine, Ronald D., "Supercomputers," ***Scientific American*** Vol. 246, No. 1, January 1982. (cited on p. 182)

[Lowry 81] Lowry, Michael R., and Allan Miller, "A General Purpose VLSI Chip for Computer Vision with Fault-tolerant Hardware," ***Proc. ARPA Image Understanding Workshop,*** April 1981. (cited on p. 184)

[Manuel 81a] Manuel, Tom, "Japanese Map Computer Domination," ***Electronics,*** Vol. 54, No. 23, November 1981. (cited on p. 183)

[Manuel 81b] Manuel, Tom, "West Wary of Japan's Computer Plan," ***Electronics,*** Vol. 54, No. 25, December 1981. (cited on p. 183)

[Marks 80] Marks, Philip, "Low-level Vision Using and Array Processor", ***Computer Graphics and Image Processing,*** Vol. 14, 1980. (cited on p. 184,185)

[Maron 81] Maron, Neil, and Thomas A. Brengle, "Integrating an Array Processor into a Scientific Computing System," ***Computer,*** Vol. 14, No. 9, September 1981. (cited on p. 183)

[S-1 79] The S-1 Project, "Fiscal Year 1979 Annual Report," UCID-18619, Lawrence Livermore National Laboratory, September 1979. (cited on p. 182,182)

[Thacker 79] Thacker, McCreight, Lampson, Sproull, Boggs, "Alto: A Personal Computer," CSL-79-11, Xerox PARC, August 1979. (cited on p. 182)

[Waller 81] Waller, Larry, "LISP Language Gets Special Machine," ***Electronics,*** Vol. 54, No. 17, August 1981. (cited on p. 183)

Chapter 7

PSYCHOLOGY AND NEUROPHYSIOLOGY

7.1 Psychology and Neurophysiology Summaries

The psychological and neurophysiological papers whose summaries follow have, in general, had significant impact upon the methods and techniques of computer vision researchers. This influence of human vision studies on machine vision work will continue to grow as psychologists and computer scientists increase the interaction of their research efforts. The categorizing that has been done on the papers has been intended to organize them by the specific aspects of vision they address. Some contribute to the field in quite broad ways and, although included in only one section, span several of the categories, while others have been included merely for the intrinsic interest of the observations they present.

7.1.1 – Parameters and mechanisms for correspondence

The papers abstracted or summarized here introduce metrics and mechanisms observed from human vision that have important effects on computer models of perception. They relate specifically to questions of the parameters used in stereo correspondence: edge detectors, relative positions, orientations, spatial frequency analysis, et cetera.

Barlow, Blakemore and Pettigrew 1967

Barlow, H.B., C. Blakemore and J. D. Pettigrew, "The neural mechanism of binocular depth discrimination," Journal of Physiology, 193, 327-342, 1967.

This is the landmark paper which introduced physiological data for disparity detectors in the visual cortex of the cat. This early paper describes convincing data from 87 cells of area 17 in 7 cats which show maximum binocular facilitation at horizontal disparities up to 6.6 degrees and vertical disparities up to 2.2 degrees. Cell 13/20, for example, which fires most effectively with a disparity between oriented bars of 3.3 degrees, did not fire at all when binocularly driven with the same optimal stimulus bars at 2.8 degrees. The cell also fired poorly when driven monocularly. Although binocularly driven cells had been studied earlier, it had been asserted that binocular stimulus always facilitated cell firing activity and that the stimulus was the same for each eye ([Hubel 62]). A possible physiological basis for stereo depth perception is substantiated by this study of binocular cells with carefully stabilized eye position. It is argued that these cells could account for solving the correspondence problem by being stimulus specific (oriented bar stimulus) and could provide the disparity determination necessary for depth perception.

Blakemore 1970

Blakemore, C., "A new kind of stereoscopic vision," Vision Research vol 10(11) 1181-1199 1970.

Researchers have sought and found binocular neurones in the cat's and monkey's visual cortex that have their receptive fields on non-corresponding points in the two retinae ([Barlow 67], [Nikara 68], [Hubel 70]). The horizontal disparity of the receptive fields varies greatly from cell to cell, and different neurones will consequently be stimulated by objects lying at different distances from the eyes. The author here suggests that this gross analysis of retinal position is not the only way in which binocular signals can be interpreted in order to retrieve the vital third dimension of visual space. He discusses an alternative approach, one based upon the spatial periodicity of the retinal images. The author presents evidence that, in certain circumstances, depth perception is related more to the spatial periodicity of the retinal images than to the actual locus of any part of those images.

Ramachandran, et al. 1973

Ramachandran, V., Rao, B. Madhusudhan, Sriram, S., Vidyasagar, T.R., "The role of colour perception and 'pattern' recognition in stereopsis," Vision Research 1973 vol 13(2) 505-509.

The authors conducted 4 experiments in which 6 trained observers viewed in a stereoscope 2 half-images consisting of random patterns of dots or rows of dots. Stereopsis was reported with simultaneous color rivalry for stimuli consisting of red and green dots. When clusters of red dots were placed in a dot matrix of green with a 3°26' disparity between red clusters, the colors provided a strong cue for stereopsis. For random dot matrices with 'linear arrays' forming square patterns, the disparity of the Gestalten provided a strong cue for stereopsis. When rows of dots with .5 and .4 cm distances were presented, results indicate that the brain preferred a regular to an irregular pattern in depth. Findings show that pattern and color recognition exert a great influence on stereopsis. Neurophysiological explanations for the results are considered.

Richards and Kaye 1974

Richards, Whitman and Kaye, Martin, "Local versus global stereopsis: Two mechanisms?," Vision Research 1974 vol 14(2) 1345-1347.

Data from the measurement of depth sensations in 3 human observers for a range of disparities and bar widths show no evidence that stereopsis involves separate local and global modes of processing.

Julesz 1974

Julesz, Bela, "Cooperative phenomena in binocular depth perception," American Scientist 1974 vol 62(1) 32-43.

The authors here interpret various stereoscopic phenomena in binocular depth perception as events within a cooperative system. Stereopsis is shown to conform to properties of a cooperative system, since various experimental results indicate the presence of disorder-order transitions, hysteresis, and multiple stable states. These stereoptical cooperative manifestations do not require form recognition, and are thus on a more primitive level than heuristic optical search procedures. It is predicted that by understanding the difference between cooperative structures and heuristic search procedures, more can be learned about the problems of semantics and intelligence.

Foley, Applebaum and Richards 1975

Foley, J.M., Applebaum, T.H., Richards, W.A., "Stereopsis with large disparities: Discrimination and depth magnitude," Vision Research 15, 417-421, 1975.

In this research, depth discrimination and perceived depth magnitude (indicated by a manual pointing response) were studied for disparities between 0.5° and 8°. For briefly exposed targets, discrimination and depth magnitude yield somewhat different functions of disparity. Discrimination begins to decline when depth magnitude is still increasing. This suggests that the variance of the depth signal increases faster than the magnitude of the signal. This result is consistent with the notion that stereroscopic processing involves the pooling of the activity of disparity detectors.

Julesz 1976

Julesz, Bela, "Global Stereopsis: Cooperative Phenomena in Stereoscopic Depth Perception," in Handbook of Sensory Physiology VIII, R. Held, H. Leibovitz and H-L. Teuber, editors, Springer, Berlin, 1976.

In this article, Julesz discusses many issues relevant to both human and mechanized correspondences of stereopsis. The theme centers around cooperative phenomena in depth perception, includes comments on the differences between global, local, coarse, and fine stereopsis, and presents his *dipole* model as an analogy to human stereopsis. This is a highly recommended article.

Frisby and Mayhew 1977

Frisby, John and Mayhew, John, "Global processes in stereopsis: some comments on Ramachandran and Nelson," Perception 1977 6(2) 195-206.

Frisby and Mayhew here discuss the use of the term 'global' in the context of stereopsis. It is concluded that different meanings of this term need to be carefully distinguished at all times. The discussion centers around a series of demonstrations introduced by V.S. Ramachandran and J.I. Nelson (1976), and interpretations are offered for these demonstrations in terms of spatial-frequency-tuned stereopsis channels.

Mayhew, Frisby and Gale 1977

Mayhew, John, Frisby, John, and Gale, Peter, "Computations of stereo disparity from rivalrous texture stereograms," Perception 1977 vol 6(2) 207-208.

This paper presents results from a computer simulation which confirm the speculation of Frisby and Mayhew that partial point-for-point correspondence can be a sufficient basis for the computation of stereo disparity by a point-for-point mechanism utilizing cooperative processes.

Nelson, Kato and Bishop 1977

Nelson, J.I., H. Kato, and P.O.Bishop, "Discrimination of orientation and position disparities by binocularly activated neurons in cat striate cortex," Journal of Neurophysiology, 40(2):260-283 1977.

The authors examined and compared the ability of binocularly activated striate cortex neurons to make both position and orientation disparity discrimination in the anesthetized and paralyzed cat. Ranges of optimal stimulus orientation disparities were found. Neurons were finely tuned to their particular orientation disparity. Since the monocular ranges of a binocular cell were not identical, orientation disparity measurement exists. Oblique orientations are less finely tuned than are verticals or horizontals, and position disparity is more finely tuned than orientation disparity.

Richards 1978

Richards, W., "Mechanisms for stereopsis," Frontiers in Visual Science 387-395 1978.

Richards here points out that stereopsis is a problem that must not be considered in isolation, but rather with respect to the questions it is designed to answer. Two such questions are: "Where are things in the three dimensional world, and what are these things?" Within these two broad categories, stereopsis is only one of several mechanisms brought into action. All of these mechanisms tend to reinforce or deny the assertions made by one another, and hence interact. Important constraints upon stereopsis are imposed by directionality, spatial frequency, contrast, and orientation. Each of the two major stereo systems may place special emphasis on one rather than another of these cues, depending upon the task at hand. An attempt is made to highlight some of these distinctions at several conceptual levels.

Mayhew and Frisby 1978

Mayhew, John and Frisby, John, "Stereopsis masking in humans is not orientationally tuned," Perception 1978 vol 7(4) 431-436.

A stereopsis signal carried by an oriented random texture and masked by a similar noise texture is not unmasked when the orientation of the noise is rotated. This experimental result is discussed in connection with the orientational tuning of local and global stereopsis processes.

Frisby and Mayhew 1978

Frisby, John, and Mayhew, John, "Contrast sensitivity function for stereopsis," Perception 1978 vol 7(4) 423-429.

Contrast thresholds for stereopsis from narrow-band filtered random-dot stereograms were compared with contrast thresholds for simple detection of similar narrow-band noise. Using the authors as observers, the study found that the contrast sensitivity function for stereopsis is similar in shape to that for detection, suggesting that as far as contrast requirements are concerned, the mechanisms of global stereopsis do not show a bias in sensitivity to any particular spatial frequency but instead require a constant level of suprathreshold contrast regardless of spatial frequency.

Mayhew, John and Frisby, John, "Contrast summation effects and stereopsis," Perception 1978 vol 7(5) 537-550.

Contrast thresholds for stereopsis were measured for a variety of bandpass- filtered random-dot stereograms in 3 experiments; 14 inexperienced subjects and the authors were the observers. The principal finding shows that contrast thresholds for stereopsis from 'complex' stereograms composed of mixtures of (a) 2 widely different spatial frequencies of (b) 2 or more widely different oriented random textures, are considerably lower than would be expected if stereopsis from such stimuli is mediated by the 1st component to rise above its own stereopsis contrast threshold. Instead, it appears that stereopsis comes about whenever the supradetection-threshold contrast of a stereogram exceeds a certain level, regardless of whether this contrast is provided by a single component or by a mix of 2 different ones. Implications for models of stereopsis are discussed.

Levinson and Blake 1979

Levinson, Eugene and Blake, Randolph, "Stereopsis by harmonic analysis," Vision Research 1979 vol 19(1) 73-78.

Stereoscopically viewed vertical gratings whose cycle widths differ by 10% fuse into a single grating rotated in depth. The perceived tilt can be predicted on the basis of a barwise computation of geometrical disparities or on the basis of a comparison in terms of harmonic content. In the present experiment, the authors found that monocular gratings of similar cycle width, but different harmonic content could not be fused, whereas gratings similar in harmonic content, but not in cycle width gave good depth sensations. Such observations support the idea of stereopsis by harmonic analysis.

Tyler and Sutter 1979

Tyler, Christopher and Sutter, Erich, "Depth from spatial frequency difference: an old kind of stereopsis?," Vision Research 1979 Vol 19(8) 859-865.

With the 2 authors as observers, perception of tilt in depth on the basis of spatial frequency difference between the 2 eyes was subjected to new tests to determine whether it could be explained by conventional binocular disparity mechanisms. When usual disparity cues were invalidated by rapidly changing displays and by using stimuli uncorrelated between the 2 eyes, perception of tilt remained for a great range of spatial frequency differences. It is suggested that the mechanism involved may be more primitive than the conventional disparity mechanism.

Blake and Cormack 1979

Blake, Randolph and Cormack, Robert, "Does contrast disparity alone generate stereopsis?," Vision Research 1979 vol 19(8) 913-915.

In 2 experiments, one using a nulling technique and the other forced-choice testing, observers judged the rotation in depth of dichoptically viewed, vertical gratings of unequal contrast. In neither experiment was evidence found for a stereoscopic sensation of depth based on contrast disparity alone; spatial frequency disparity, on the other hand, produced a clear rotation of the fused percept about the vertical axis.

Mayhew 1980

Mayhew, John, "The computation of binocular edges," Perception 1980, vol 9(1) 69-86.

Mayhew here describes a computational model that effects the binocular combination of monocular edge information. The distinctive features of the model are that it: (a) identifies edge locations in each monocular field by searching for zero crossings in nonoriented center-surround convolution profiles, (b) selects among all possible binocular point-for-point combinations of edge locations only those which satisfy a (quasi-)collinear figural grouping rule, and (c) presents a concept of the oriented and spatial-frequency- tuned channel as a nonlinear grouping operator. The success of the model is demonstrated both on a stereo pair of a natural scene and on a random- dot stereogram.

Burt and Julesz 1980

Burt, P. and Julesz, B., "A disparity gradient limit for binocular fusion," Science vol 208(4444) 615-617, 1980.

Ever since Panum, it has been commonly assumed that there is an absolute disparity limit for binocular fusion. Burt and Julesz report here that nearby objects modify this disparity limit. This result sheds new light on several enigmatic phenomena in stereopsis. Contrary to the common assumption (Panum) that disparity magnitude is the limiting factor for fusion, when two or more objects occur near one another in the visual field, it is found that disparity gradient is the limiting factor. Fusion of at least one object fails when the gradient exceeds a critical value of approximately unity.

7.1.2 - Spatial frequency filtering and spatial pooling

One of the most pervasive influences from psychology and neurophysiology research has been the introduction of spatial frequency filtering for feature detection. The evidence from these studies suggests that the human visual system has four (or five) spatial-frequency-tuned filtering mechanisms operating on the intensity data in the visual field. Most of the debate centers on the shape and bandwidth of these filtering mechanisms, and the relationships between activity in the various bands.

Julesz 1975

Julesz, Bela, "Two-dimensional spatial-frequency-tuned channels in visual perception," Proc. of Intern. Conf. on Signal Analysis and Pattern Recognition in Biomedical Engineering 177-197 1975.

Julesz and Miller 1975

Julesz, Bela and Miller, Joan, "Independent spatial-frequency-tuned channels in binocular fusion and rivalry," Perception 1975 vol 4(2) 125-143.

In the experiments described in this paper, random-dot stereograms presented to subjects were bandpass filtered in the 2-dimensional Fourier domain, and masking noise of various spatial frequency bands was added to the filtered stereograms. Masking noise bands containing equally effective noise energy were selected such that their bands were either overlapping with the stereoscopic image spectrum or were 2 octaves distant. The first case resulted in binocular rivalry; however, in the second case stereoscopic fusion could be maintained in the presence of strong binocular rivalry owing to the masking noise. This finding indicates that spatial-frequency-tuned channels are not restricted to 1-dimensional gratings but operate on 2-dimensional patterns as well. Furthermore, these frequency channels are utilized in stereopsis and work independently from each other, since some of these channels can be in binocular rivalry, while at the same time other channels yield fusion.

Wilson and Giese 1977

Wilson, H.R., Gięse, S.C., "Threshold visibility of frequency gradient patterns," Vision Res. 17, 1177-1190, 1977.

In this work, the role of spatial inhomogeneity in threshold grating peception was studied using grating patterns containing a gradient of spatial frequencies. Based on both psychophysical and neurophysiological evidence, it was decided that an appropriate class of patterns would be those with a linearly varying spatial wavelength. Thresholds were measured both for patterns in which the spatial wavelength increased linearly with eccentricity and for patterns in which it decreased linearly with eccentricity. The results demonstrated that spatial inhomogeneity is indeed an important factor in the threshold visibility of gratings. The data support medium bandwidth estimates for the mechanisms underlying spatial frequency selectivity, and they are , inconsistent with the notion that the visual system performs a Fourier analysis of visual images. The data can be fit quantitatively with a semi-empirical model which postulates that the sensitivity to all spatial frequencies is highest in the fovea, but that the sensitivity to high spatial frequencies declines more rapidly than that to low frequencies with increasing eccentricity. A comparison of the model with measured line spread functions permits some estimates to be made of the range of receptive field sizes present at each eccentricity.

Wilson and Bergen 1978

Wilson, Hugh R., James R. Bergen, "A four mecahanism model for threshold spatial vision," University of Chicago, April 1978.

Data on the threshold visibility of spatially localized, aperiodic patterns were used to derive the properties of a general model for threshold spatial vision. The model consisted of four different size-tuned mechanisms centred at each eccentricity, each with a centre surround sensitivity profile described by the difference of two Gaussian functions. The two smaller functions showed relatively sustained temporal characteristics, while the larger two exhibited transient properties. All four mechanisms increased linearly in size with eccentricity. Mechanism responses were combined through spatial probability summation to predict visual thresholds. The model quantitatively predicts the spatial modulation transfer function under both sustained and transient conditions with no free parameters.

Schumer 1979

Schumer, Robert, "Independent stereoscopic channels for different extents of spatial pooling," Vision Research 1979 Vol 19(12) 1303-1314.

Schumer conducted 2 experiments on 2 adult human subjects, with stimuli composed of dynamic visual noise stereograms in which binocular disparity was modulated sinusoidally with changes in vertical spatial position. The first experiment showed that observers can just detect a compound disparity grating when at least 1 of 2 presented sinusoidal components is close to its own independent threshold amplitude. The second experiment demonstrated selective threshold elevation following prolonged viewing of a disparity grating (selective adaptation). Results demonstrated that the human visual system contains multiple, independent stereoscopic mechanisms selectively tuned for different spatial frequencies of disparity modulation, each characterized by a different extent of spatial pooling. The observed bandpass characteristic implies that these mechanisms must receive lateral inhibition from disparity detectors tuned to adjacent positions in space.

7.1.3 – Isoluminance studies

A particularly fascinating observation from human perceptual studies is that arising from isoluminance experiments. Here, patterns of differing colors but identical luminance have been used to demonstrate the apparent existence in humans of two separate mechanism for stereopsis. One allows the attainment of fusion of random-dot stereograms, but fails to function when the dots are of equivalent brightness. The other apparently processes color information in a more global way, and provides depth perception where there are extended monocular features (*contours*) despite isoluminance across the features.

Lu and Fender 1971

Lu, C. and Fender, D., "Interaction of color and luminance in stereoscopic vision," 1971 Annual Meeting of the Optical Society of America.

Lu and fender describe here experiments conducted using the random-dot stereo patterns devised by Julesz, but substituting various colors and luminances for the usual black and white random squares. They found that the ability to perceive the patterns in depth depends on a luminance difference between the colors used. If two colors are the same luminance, then depth is not perceived although each of the individual squares which make up the patterns is easily seen because of the color difference. This was found true, for all tested combinations of different colors. If different colors are used for corresponding random squares between the left- and right-eye patterns, stereopsis is possible for all combinations of binocular rivalry in color, provided the luminance always precludes stereopsis, regardless of the colors involved.

Gregory 1977

Gregory, Richard, "Vision with isoluminant colour contrast: 1. a projection technique and observations," Perception, 6, 113, 1977.

An optical technique is described here for projecting two-colour pictures with controlled brightness contrast, which may be set to zero for isoluminance. Colour registration is maintained without adjustment or special setting up. It is suggested that colour- and brightness- contour registration in the visual channel is a problem which may be solved neurally by master brightness signals locking slave colour signals. The projection apparatus allows the supposed master brightness signals to be removed - at isoluminance - when contour disturbances should occur. The observation, originally reported by Lu and Fender [Lu 71] and confirmed here, that random-dot stereograms lose depth with isoluminance, suggests that this kind of stereo by cross-correlation of points is dependent on the luminance channel. It also suggests (though it does not definitively show) that stereo depth by random dots is given by a different mechanism from stereo by contours, since ***contour*** stereo is not lost with isoluminance though random-dot stereo (with its illusory contours) is lost.

de Weert 1979

de Weert, Charles, "Colour contours and stereopsis," Vision Research 1979 vol 19(5) 555-564.

De Weert describes studies using 2 observers to investigate the role of color contours in stereopsis for random dot stereograms and for figural stereo stimuli. The main conclusion is that for random dot stereograms there were ratios of luminances for dots and background where stereopsis disappeared. This points to the possibility that color differences alone are not sufficient to evoke stereopsis. The ratio of the luminances, however, was not unity when dots and background were of different colors. This points to color-specific effects. For figural stereograms, stereopsis did not disappear at the aforementioned ratios of the luminances of figure and background. Here, color certainly had a contribution to stereopsis on its own. In another series of experiments the influence of color rivalry on the stereo thresholds was investigated. Results suggest that the color information used in the figural system to evoke stereopsis is the same for both eyes.

Russell 1979

Russell, P.W., "Chromatic input to stereopsis," Vision Research 19(7) 831-834 1979.

Lu and Fender ([Lu 71]) presented Julesz random dot patterns in which the components of the array directed to each eye differed from each other in color but not in luminance. They found that a disparity between the images presented to the two eyes in this way gave no impression of depth. The present authors show that the results of Lu and Fender may be somewhat better explained by postulating that the perception of depth under their conditions depends on the R-G channel of conventional opponent colors theory.

7.1.4 - Monocular cues to stereopsis

It is clear that humans can 'see' the solidness of their world through either binocular viewing (simultaneous stereopsis) or temporal viewing (perhaps a moving monocular observer or a changing scene). The interaction between this monocular and binocular functioning is being studied extensively in perceptual psychology. This is generally done through latency experiments, where improvements in response time (i.e. the attainment of stereopsis) arising from the introduction of structural features to the field of view (which generally contain some form of random dot stereogram) suggest the increased performance attainable when the two mechanisms work together. There is an obvious relationship between these studies of monocular facilitation of stereopsis and the previous section's concern with *edge* versus *contour* (or *local* versus *global*) stereopsis mechanisms.

Saye and Frisby 1975

Saye, Ann and Frisby, John, "The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms," Perception 1975 vol 4(2) 159-171.

In this work, 2 experiments with a total of 42 normally seeing undergraduates investigated the effects on stereopsis perception times of including monocularly conspicuous features in random-dot stereograms. It was found that such features facilitated stereopsis in large-disparity but not in small-disparity stereograms, perception times for the latter being relatively short with or without monocular features. Facilitation in the large-disparity stimuli came about both from features which delineated the shape of the whole disparate area and from features which merely happened to lie in the same depth plane as the disparate area, but which did not give any shape cues. It is argued that these various results can be well accounted for by a 'vergence hypothesis', which supposes that the long perception times often found with random-dot stereograms are due in part to the absence of stimulus features which can guide the vergence movements necessary for fusing the display.

Saye 1976

Saye, Ann, "Facilitation of stereopsis from a large disparity random-dot stereogram by various monocular features: Further findings (A short note)," Perception 1976 5(4) 461-465.

The author examined the effects of adding 5 different kinds of prominent monocular features to a large-disparity random-dot stereogram. 30 university students, randomly assigned to 1 of 5 groups, received 15 presentations of 1 of 5 kinds of stimuli used, followed by 3 presentations of a small-disparity demonstration stimulus. Latency for fusion of each stimulus presentation was recorded. It was found that features which enclosed the disparate area produced the shortest initial perception times for fusion. The longer initial perception times for stimuli containing features without this enclosing property are explained in terms of less-helpful guidance of saccadic eye movements prior to the establishment of fusion. Subsequent reductions in perception times for these latter stimuli may be due to perceptual learning within the eye movement control system.

Richards 1977

Richards, W, "Stereopsis with and without monocular contours," Vision Research vol 17(8) 967-9 1977.

The depth perception elicited in 4 subjects from random dot stereograms devoid of monocular cues was severely impaired when compared with similar stereograms that revealed the monocular contours. For transient stimuli, monocular contours appear necessary to elicit a range of depth sensations for different disparities, suggesting that monocular cue analysis is an integral component of the stereomechanisms.

Julesz 1978

Julesz, Bela, "Binocular utilization of monocular cues that are undetectable monocularly," Perception 1978 vol 7(3) 315-322.

The latency of tracking dynamic random-dot stereograms can be shortened by as much as 100 msec when monocular cues are added by introducing a difference in dot density between target and surround. It has been tacitly assumed that perception time will be reduced only if the added monocular cues are above the detection threshold for each eye. However, the experiments reported here clearly show that stereoscopic performance as measured by an eye tracking task can be greatly enhanced by added monocular cues that cannot be detected. Two observers were instructed to track a suddenly displaced vertical bar (portrayed as a dynamic random-dot stereogram) while their eye movements were recorded by EDG. The bar had either a given binocular disparity or zero binocular disparity with respect to its surround. For the target with a disparity (in a wide range), the latency time of tracking decreased by more than 30 msec (10%) as density differences increased from 0 to 4%, whereas in the control conditions with no stereoscopic cues (zero disparity) subjects were unable to track the bar at all within that range of density difference. Thus stereopsis is greatly aided by minimal monocular cues that by themselves elude monocular detection.

Yonas, Cleaves and Pettersen 1978

Yonas, A., Wallace T. Cleaves, and Linda Pettersen, "Development of Sensitivity to Pictorial Depth," Science, Vol. 200, 77-79, April 1978.

Sensitivity to static pictorial information for depth develops between 22 and 26 weeks of age. When conflicting binocular and surface-texture information was minimized, 26 to 30 week old infants directed their reaching to the apparently closer side of a photograph of a window rotated in depth. Younger infants, from 20 to 22 weeks of age, did not direct their reaching to the pictorially nearer side of the display, but did reach with a high degree of directionally when presented with a real window rotated in depth.

Kidd Frisby and Mayhew 1979

Kidd, A.L., Frisby, J.P., Mayhew, J.E.W., "Texture contours can facilitate stereopsis by initiating vergence eye movements," Nature vol 280(5725) 829-832 1979.

The authors show here that vergence movements are not always random in the stimulus circumstances of textured contours. Rather, vergence can be guided by texture contours depicted by differences of texture orientation between regions of relatively high spatial frequency content for the disparity range incorporated in the stereogram.

7.1.5 – Stereo acuity measures

The following papers highlight the research being done in determining the factors affecting perceptual acuity.

Richards and Foley 1974

Richards, Whitman and Foley, John, "Effect of luminance and contrast on processing large disparities," J. of the Optical Soc. of America 1974 vol 64(12) 1703-1705.

This paper describes a study conducted by the authors, with themselves as observers. They found that although reduced luminance impaired the discrimination of small disparity stimuli, large disparity discrimination improved. Crossed and uncrossed stimulus disparities of 4° that were not discriminated at photopic levels were easily discriminated at mesopic levels near the color threshold. This improvement of stereo processing appeared to be dependent upon an effective contrast reduction produced neurally, because a physical reduction of contrast without a change of background luminance also improved large-disparity stereopsis.

Herman, Tauber and Roffwarg 1974

Herman, John, Edward Tauber, Howard Roffwarg, "Monocular occlusion impairs stereoscopic acuity, but total visual deprivation does not," Perception and Psychophysics 1974 vol 16(2) 225-228.

The authors designed an apparatus for testing stereoscopic accuracy which eliminated all cues to depth except binocular disparity. The relative effect of 8 hrs of monocular- as opposed to binocular-occlusion on subsequent stereoscopic performance was tested in 6 subjects. Monocular patching led to significant increases in mean standard deviation and in mean absolute error as compared to baseline testing. Binocular patching led to no such impairment. Thus, true disuse (such as occurs during binocular deprivation) did not impair stereopsis, whereas monocular occlusion, which may involve temporary misuse of the stereoscopic system, did.

Uttal, Fitzgerald and Eskin 1975

Uttal, William, Judy Fitzgerald, Thelma Eskin, "Rotation and translation effects on stereoscopic acuity," Vision Research 1975 vol 15 (8-9) 939-944.

The authors examined the effects of rotation or translation on the detectability of a dotted target plane embedded in a dotted masking cube. Julesz-type stereograms were generated in real time, using a small laboratory computer. Three students displayed an excellent ability to compensate for x-y projected density variations produced by changes in the angle of rotation of the target plane so that performance was equal at all rotations. This finding indicates that dot density is being processed in a 3- rather than a 2-dimensional manner. Translation of a frontoparallel plane from the back of the stereoscopically defined cube to its front produced systematic changes in performance. Performance was best at the central fixation distance and decreased as either crossed or uncrossed disparity increased. Thus, it appears that stereo acuity decreases symmetrically with increases in disparity in either direction.

7.1.6 – Diverse observations

The following papers are a sampling of some diverse but interesting aspects of human stereopsis.

Mitchell and Blakemore 1970

Mitchell, D.E. and Blakemore, C., "Binocular depth perception and the corpus callosum," Vision Research vol 10(1) 49-54 Jan 1970.

The authors here point out that an object lying directly behind or in front of the fixation point has images that project to separate hemispheres through the two eyes. A split-brain human cannot interpret the depth of such an object although his peripheral stereopsis is normal. They conclude that there must be an interhemisphere link for binocular integration in central vision.

Movshon, Chambers and Blakemore 1972

Movshon, J.A., Chambers, B.E., Blakemore, C., "Interocular transfer in normal humans and those who lack stereopsis," Perception 1972 vol 1(4) 483-490.

This paper described the investigation of interocular transfer of the tilt aftereffect in 15 University students: 8 subjects with good stereopsis and 7 subjects without stereoscopic vision. The latter subjects were divided into 2 groups: 4 with and 3 without a history of strabismus. Strabismic subjects showed grossly reduced interocular transfer of the effect (12% mean transfer). Nonstrabismic subjects had moderate transfer (49%) and normal subjects showed approximately 70% mean transfer. All normal subjects showed greater transfer from the dominant eye to the nondominant than vice versa. Results are discussed with respect to developmental effects in the visual system of cats and humans, and the nature of the tilt aftereffect.

Mitchell and Baker 1973

Mitchell, Donald and Baker, Andrew, "Stereoscopic aftereffects: Evidence for disparity-specific neurones in the human visual system," Vision Research 1973 vol 13(12) 2273-2288.

Mitchell and Baker found, with 2 human subjects that following adaptation to a target imaged with a certain disparity, the apparent depth of targets imaged with nearby disparities was altered. With simple stimuli (single lines), the maximum displacement occurred with adapting disparities of about 5', but adapting disparities exceeding 15-20' had no effect. On the other hand, following adaptation to a vertical grating, the apparent depth of a test grating was altered in a cyclical fashion with increasing adapting disparity, being displaced first in one direction and then the opposite with a period equal to twice the spatial period of the grating stimulus itself. Such periodic adaptation curves were found only if both test and adapting gratings were vertical. Results are compared with the properties of disparity- specific neurones in the cat and monkey visual cortex.

Regan and Beverley 1973

Regan, D. and Beverley, I., "Electrophysiological evidence for existence of neurones sensitive to direction of depth movement," Nature 1973 vol 246(5434) 504-506.

This report discusses a series of 15 experiments which investigated human visual and electrical brain responses to stereoscopic and monocular stimuli. Subject's right and left eyes viewed a 5° pattern of randomly arranged black dots. When the left eye alone was used, subject saw a 2° section move from side to side (monocular condition) and when the 2 patterns were viewed by both eyes, the movement produced an illusion that the 2° section was moving in and out (stereoscopic condition). In some experiments, a bar stimulus was used so that the left eye's retinal image of the bar could be moved. Stereoscopic stimulation produced very different responses than monocular stimulation for both the bar and dot patterns. Results indicate that the way in which the brain handles information that a target's retinal disparity has changed depends on whether the target is in front of or behind the fixation point. Information of a change in retinal disparity is perceived different if the target is moving away rather than toward the fixation point. It is suggested that the different responses made to depth movements reflect the activities of neurones selectively responsive to the movement of a target whose disparity is changing.

Harris and Gregory 1973

Harris, J.P. and Gregory, R.L., "Fusion and rivalry of illusory contours," Perception 1973 vol 2(2) 235-247.

The authors question whether visual contours are given directly from striate-cortex feature-detector activity. Phenomena of 'subjective' or 'cognitive' contours are presented to examine this view, on the ground that contours can be extrapolations across low-probability gaps. The contours may be curved and may have poor 'gestalt' qualities. Therefore, 'gestalt closure' is not appropriate, but may be a subclass of these phenomena. It is suggested that these illusory contours (and brightness differences) are generated by perceptually postulated masking objects which are part of perceptual 'scene analysis strategy,' since strong evidence for nearer objects is provided by improbable gaps. Experiments with 10 subjects are reported in which each eye was given a different 'cognitive' contour figure such that there were disparate but illusory contours. It was found that these were fused to give 3-dimensional illusory surfaces bowing in front of the display. Results indicate that masking objects must be in front of gaps, and that switching the eyes often gives rivalry of the illusory contours when masking is incompatible with the stereo depth. Implications for normal stereo vision are discussed.

Luria and Kinney 1975

Luria, S. and Kinney, J., "Vision in the water without a facemask," Aviation, Space and Environmental Medicine 1975 vol 46.

Distance and size estimates and stereoacuity judgments were made in water by divers, both with and without facemasks. Without the mask, only stereoacuity was markedly degraded. Distance estimates were slightly more accurate, despite a great decrease in the range of visibility. Size estimates were slightly too small. Subjects with refractive errors did not appear to be more handicapped than subjects with normal vision.

Rogers and Anstis 1975

Rogers, Brian and Anstis, Stuart, "Reversed depth from positive and negative stereograms," Perception 1975 vol 4(2) 193-201.

A stereogram was presented with patterns of opposite contrast – one positive the other negative. One eye received only the positive pattern; the other received both positive and negative patterns superimposed. Subjects reported apparent reversals of perceived depth: crossed (convergent) disparity made the fused stereogram appear further away. While uncrossed (divergent) disparity made it appear nearer. It is suggested that spatial summation in the visual system blurred the superimposed positive-and-negative contours and shifted their effective positions, leading to reversals in perceived disparity.

Frisby and Julesz 1975

Frisby, John and Julesz, Bela, "The effect of orientation differences on stereopsis as a function of line length," Perception 1975 vol 4(2) 179-186.

Experimental results with 2 highly practiced subjects show that the amount of depth seen in a multiline stereogram composed of horizontal lines steadily decreased as the lines in one field of view were rotated about their midpoints. This effect of orientation difference on stereopsis was more acute the longer the lines in the stereogram. It is suggested that the critical factor underlying the depth reduction is not orientation difference per se, but rather the vertical disparity which an orientation difference introduces into the display between the tips of corresponding lines. This interpretation is supported by the fact that similar vertical disparities caused similar depth reductions regardless of the length of the lines in the stereogram.

Anstis and Rogers 1975

Anstis, S. and Rogers, B., "Illusory reversal of visual depth and movement during changes of contrast," Vision Research vol 15(8-9), 957-961, 1975.

The visual system usually sees apparent movement (phi) when two similar pictures are exposed successively, and stereoscopic depth when the pictures are exposed one to each eye. But when a picture was followed via a dissolve by its own photographic negative, overlapping but displaced, strong apparent movement was seen in the opposite direction to the image displacement ('reversed phi'). When both eyes saw a positive picture, and one eye also saw an overlapping low-contrast negative containing binocular disparity, 'reversed stereo' was seen, with the apparent depth opposite to the physical disparity. Results were explained with a model of spatial summation by visual receptive fields.

Kaufman 1976

Kaufman, Lloyd, "On stereopsis with double images," Psychologia: An International Journal of Psychology in the Orient 1976 vol 19(4) 224-233.

Stereopsis may occur in the presence of large disparities despite the fact that double images are visible. Many theorists have proposed that neural fusion of the diplopic half-images produces the depth effect even though the neural fusion is not accompanied by phenomenal fusion. The paper offers the alternative explanation that stimuli containing widely disparate half-images are compound, Panum-limiting, case stereograms. It is shown that depth may be controlled by controlling disparity of fixation in both simple and compound limiting case stereograms. It is concluded that widely disparate half-images do not interact in the computation of depth.

Koenderink and Van Doorn 1976

Koenderink, J.J., and Van Doorn, A.J., "Geometry of binocular vision and a model for stereopsis," Biol. Cybern. (Germany) vol 21(1) 29-35 1976.

If a binocular observer looks at surfaces, the disparity is a continuous vector field defined on the manifold of cyclopean visual directions. This field is derived for the general case that the observer is presented with a curved surface and fixates an arbitrary point and the disparity field is expanded in the neighborhood of a visual direction. The first order approximation can be decomposed into congruences, similarities and deformations. The theory provides a geometric explanation of the percepts obtained with uniform and oblique meridional aniseikonia. The authors utilize the geometric theory to construct a mechanistic model of stereopsis that obviates the need for internal zooming mechanisms, but nevertheless is insensitive to differential cyclotorsion or uniform aniseikonia.

Hogben, Julesz and Ross 1976

Hogben, J.H., Bela Julesz, and John Ross, "Short-term memory for symmetry," Vision Research 1976 vol 16(8) 861-866.

In a study with 2 experienced (in stereopsis studies) observers, symmetric cascades of dots were generated in a continuous random sequence such that each dot had a partner reflected about a vertical or horizontal axis, respectively. Between each point and its partner a temporal delay was introduced. While the brightness of the dots appeared constant within 120 to 140 msec, symmetry perception ceased at delays in the range of 50 to 90 msec, depending on observers, type of symmetry, and plotting rate. These findings, in conjunction with 3 control studies, suggest that memory span for position information is limited to 50 to 90 msec while memory for brightness information lasts for 120 to 140 msec. Perturbation experiments (with no delay) in which a certain proportion of dots had no symmetrical partner were compared with the delay experiments for equal performance, and equivalence curves between delay and perturbation rate were obtained. While performance depended on the type of symmetry and plotting rate, the shape of equivalence curves remained unchanged.

Wallach and Bacon 1976

Wallach, Hans and Bacon, Joshua, "Two forms of retinal disparity," Perception and Psychophysics 1976 May vol 19(5) 375-382.

In 6 experiments with 147 undergraduates, the retinal disparities in stereograms where the vertical alignment of pairs of homologous points in one eye differs from that in the other eye were found to be more effective than disparities that do not involve that kind of binocular difference. The presence of such 'transverse disparities' shortened the time elapsed until perceived depth was reported in 4 instances, in 2 simple stereogram pairs and in 2 different pairs of random dot pattern stereograms. In an experiment where binocular parallax was in conflict with an effect of past experience, the presence of transverse disparities caused binocular parallax to prevail. The presumption that the amount of perceived depth depends only on the amount of disparity (provided distances from the eyes are unchanged) and not on the configuration in which it manifests itself, did not hold in stereograms containing transverse disparities.

Frisby and Julesz 1976

Frisby, John and Julesz, Bela, "The effect of length differences between corresponding lines on stereopsis from single and multi-line stimuli," Vision Research 1976 vol 16(1) 83-87.

The amount of depth seen by 2 practiced subjects in a random line stereogram composed of orthogonal corresponding lines steadily diminished as line length increased. The line length required to destroy stereopsis completely was the same regardless of the disparity incorporated in the stereogram. Results with multiline stimuli are discussed in connection with previous observations made with single-line stimuli. It is suggested that depth shift effects found with single lines can be explained by supposing that orthogonal line percepts fuse at their endpoints and that as line length is increased, this results in an altered retinal disparity of the endpoints and hence a shift in apparent depth. The findings with multiline stimuli cannot be explained in this simple fashion and require a model of stereopsis which takes into account interactions between elements in complex displays.

Smith 1977

Smith, Babington, "A wartime anticipation of random-dot stereograms," Perception 1977 vol 6(2) 233-234.

This article presents a stereoscopic pair of photographs taken over Cologne, Germany, in 1940 by a Royal Air Force Spitfire. The photos, dubbed 'The Empty Rhine', are relevant to the controversy which led to Julesz's work ([Julesz 64]) and random-dot stereograms.

Klein 1977

Klein, Raymond, "Stereopsis and the representation of space," Perception 1977 vol 6(3) 327-332.

Klein discusses tests of 4 stereoblind (as assessed by tests of subject's ability to process static and dynamic disparity cues) and 4 normal subjects on a mental rotation task. It was hypothesized that if stereopsis is an important input for building up the perceptual system that represents 3-dimensional space, then subjects lacking it ought to be deficient at mental rotations in depth. Stereoblind subjects were equally efficient at picture-plane and depth rotations and were nonsignificantly better than normal subjects at rotations in depth. It is concluded that in the absence of stereopsis other cues are sufficient for the development of the 3-dimensional perceptual system. A paradox was raised, however, by the finding that the introspections of the 2 groups differed markedly.

Lema and Blake 1977

Lema, Sandra and Blake, Randolph, "Binocular summation in normal and stereoblind humans," Vision Research 1977 vol 17(6) 691-695.

The experiments here measured monocular and binocular contrast thresholds over a range of spatial frequencies in 4 normal and 4 stereoblind 21 to 36 year old observers. Results indicate that unlike the normal subjects, whose binocular thresholds were consistently lower than monocular, stereoblind subjects were no better with 2 eyes than with one. These results suggest that stereopsis and binocular summation are mediated by a common neural mechanism.

Fox, Lehmkuhle and Bush 1977

Fox, Robert, Stephen Lehmkuhle, and Robert Bush, "Stereopsis in the falcon," Science 1977 vol 197(4298) 79-81.

This paper reports a behavioral demonstration of stereopsis in the falcon, a nonmammalian with binocular vision. An American kestrel (Falco sparverius) was trained to select a stereoscopic form in a 2-choice discrimination task. The finding of stereopsis in this species complements recent physiological evidence for binocular interaction in the bird visual system, and suggests that stereopsis may be a general attribute of vertebrate vision and not an exclusive product of mammalian evolution.

Kulikowski 1978

Kulikowski, J., "Limit of single vision in stereopsis depends on contour sharpness," Nature 1978 vol 275(5676) 126-127.

The human brain combines the disparate retinal images from both eyes so that an object is seen as single within a certain limit of disparity; this limit is called Panum's fusional area. For fine-line targets, the disparity limits have been determined to be ±7 min of arc for foveal vision and larger for peripheral vision. The experiments described here (15 normal observers) show that Panum's area increases for stimuli with blurred contours (i.e., for contours containing low spatial frequencies). Results suggest that the range of single vision in stereopsis is inversely related to the spatial frequencies present in the pattern, as though there were many Panum's areas instead of one determined with fine lines.

Hamsher 1978

Hamsher, K., "Stereopsis and the perception of anomalous contours," Neuropsychologia 1978 vol 16(4) 453-459.

A relationship between depth perception and the perception of anomalous contours has been suggested by some investigators. To test this hypothesis, performance of 50 hospital control patients from general and orthopedic wards and 23 brain-damaged patients on an anomalous contour test was compared to performance on 2 stereoscopic tests: a conventional stereoacuity test (local stereopsis and a test of random-letter stereopsis (global stereopsis) were given to patients with unilateral right hemisphere lesions, since only such patients have been shown to be specifically impaired on the latter type of test. Findings demonstrate a significant relationship between both types of stereoscopic performances and anomalous contour perception. Implications for a theory of global stereopsis are discussed.

Butler and Westheimer 1978

Butler, Thomas and Westheimer, Gerald, "Interference with stereoscopic acuity: spatial, temporal, and disparity tuning," Vision research 1978 vol 18(10) 1387-1392.

Stereoscopic acuity in 3 normal human subjects was conspicuously reduced by the presence of contours contiguous to the test pattern. Flanking contours interfered maximally when they were placed about 2.5 min from a test line and less when this distance was increased or decreased. The largest interference effect was obtained when the flanks were presented 100 msec after the onset of the test pattern. The interference had a narrow depth tuning: to halve the threshold elevation the flanks need be presented only 12, 17, and 40 sec out of fixation plane for the 3 subjects. Because a companion experiment revealed no such disparity specifically for interference with vernier acuity, the effects described here must operate in the stereo domain.

Cogan 1978

Cogan, Alexander, "Qualitative observations in visual science: 'The Farnsworth˙Shelf:' Fusion at the site of the 'ghosts'," Vision Research 1978 vol 18(6) 657-664.

The Keplerian projection theory of stereopsis predicts hidden intersections (i.e. 'ghosts') in the binocular visual space. Three experiments were performed in real and stereoscopic space with a total of 8 experienced and 9 naive subjects in experiments one and two, and 11 experienced subjects in experiment three. Results indicate that the fusional vergence may rest at the site of the 'ghosts'; this normally happens when stereograms are perfectly fused. It is suggested that fusion always acts to minimize global disparity differences, unless the fixation reflex is elicited to oppose global fusion. The author suggests that the hidden (i.e. from awareness) structure of the binocular projection field represents a valuable 'matrix' containing ready programs for changes in vergence and for perception of spatial positions. The problem of the 'ghosts' is interpreted as a normal instance of the still unknown relationship between stimulation and perception.

Friedman, Kaye and Richards 1978

Friedman, Rhonda, Martin Kaye, Whitman Richards "Effect of vertical disparity upon stereoscopic depth," Vision Research 1978 Vol 18(3) 351-352.

In a study designed to measure the depth elicited from a stimulus with a fixed binocular disparity but with the orientation of the disparity difference ranging from horizontal to vertical, data from 3 subjects were compared with the following 3 models: depth corresponds to the horizontal disparity component; a given binocular stimulus may elicit both sensations of 'pure depth' and 'pure rivalry'; and vertical disparity attenuates horizontal disparity processing. It is suggested that there may be instances in which the visual system needs to process binocular parallax in the presence of vertical disparity.

Yellott and Kaiwi 1979

Yellott, John and Kaiwi, Jerry, "Depth inversion despite stereopsis: the appearance of random-dot stereograms on surfaces seen in reverse perspective," Perception 1979 vol 8(2) 135-142.

Inside-out relief masks of faces can be depth-inverted (i.e. seen in reverse perspective) during close-up binocular viewing. If a random-dot stereogram is projected onto such a mask, stereopsis can be achieved for the stereogram, and its depth planes are correctly seen while the mask itself, including the region covered by the stereogram, is simultaneously perceived as depth-inverted. The present paper shows that binocular depth inversion cannot be explained by a complete loss of stereoscopic information (e.g., through monocular suppression), or by a process analogous to pseudoscopic viewing whereby retinal disparities are incorporated into perception but with their signs uniformly reversed.

Frisby and Mayhew 1979

Frisby, John and Mayhew, John, "Depth inversion in random-dot stereogram," Perception 1979 Vol 8(4) 397-399.

In this paper, Frisby and Mayhew contend that depth inversion is possible for a random-dot stereogram. The conditions under which the inversion can be obtained are described, and evidence is presented in support of an explanatory hypothesis.

Mayhew and Frisby 1979

Mayhew, John and Frisby, John, "Surfaces with steep variations in depth pose difficulties for orientationally tuned disparity filters," Perception 1979, vol 8(6) 691-698.

This paper discusses the difficulties for orientationally tuned disparity filters in connection with a random-dot stereogram depicting a surface with steep horizontal corrugations. Five adults previously experienced in stereopsis experiments served as subjects. As expected on theoretical grounds, it was found that a vertical ±45° orientationally filtered version of this stereogram cannot be fused. Moreover, it is demonstrated that a horizontal ±45° filtered version can be fused only with difficulty, and its stereo percept is poor compared to that of the unfiltered original. It is concluded that oriented filters seem ill-designed to mediate the extraction of disparity cues, at least in the cases under consideration.

Aslin, Dumais, Fox and Shea 1979

Aslin, R.N., Dumais, S.T., Fox, R., Shea, S.L., "Stereopsis in human infants," Science vol 207(4428) 323-324 1980.

Stereoscopic depth perception was tested in human infants by a new method based on attracting the infant's attention through movement of a stereoscopic contour formed from a dynamic random-element stereogram. The results reveal that stereopsis emerges at 3.5 to 6 months of age, an outcome consistent with evidence for rapid postnatal development of the visual system.

7.2 References

[Barlow 67] Barlow, H.B., C. Blakemore, J.D. Pettigrew, "The neural mechanism of binocular depth discrimination," *Journal of Physiology*, 193, 327-342, 1967. (cited on p. 197)

[Hubel 62] Hubel, D.H. and T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, 160, 385-398, 1962. (cited on p. 196)

[Hubel 70] Hubel, D.H. and T.N. Wiesel, "Cells sensitive to binocular depth in area 18 of the macaque monkey cortex," *Nature*, 225, 41-42, 1970. (cited on p. 197)

[Julesz 64] Julesz, Bela, "Binocular depth perception without familiarity cues," *Science*, 145(3630) 1964. (cited on p. 214)

[Lu 71] Lu, C. and Fender, D., "Interaction of color and luminance in stereoscopic vision," 1971 Annual Meeting of the Optical Society of America. (cited on p. 205)

[Nikara 68] Nikara, T., P.O. Bishop, J. Pettigrew, "Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex," *Exp. Brain Research*, 6, 353-372, 1968. (cited on p. 197)

Chapter 8

BIBLIOGRAPHY

[Abdou 1978] Abdou,I.E., "Quantitative Methods of Edge Detection," Ph.D. thesis, University of Southern California, July 1978. Also USCIPI Report 830.

[Ackerman 1982] Ackerman, William B., "Data Flow Languages," *Computer,* Vol. 15, No. 2, February 1982.

[Allam 1978] Allam, M.M., "DTM Applications in Topographic Mapping," *Photogrammetric Engineering and Remote Sensing,* vol. 44, no. 12, 1513, December 1978.

[Altes] Altes, R.A., "Spline-like Image Analysis with a Complexity Constraint. Similarities to Human Vision," unpublished paper, ca. 1975, 36 pp.

[Anstis 1975] Anstis, S. and Rogers, B., "Illusory reversal of visual depth and movement during changes of contrast," *Vision Research,* Vol. 15(8-9), 957-961, 1975.

[Arnold 1978] Arnold, R. David, "Local Context in Matching Edges for Stereo Vision," *Proceedings of the ARPA Image Understanding Workshop,* Boston, 65-72, May 1978.

[Arnold 1980] Arnold, R.D., and T.O. Binford, "Geometric Constraints in Stereo Vision," *Soc. Photo-Optical Instr. Engineers,* vol. 238, Image Processing for Missile Guidance, 281-292, 1980.

[Arnold 1982] Arnold, R. David, "Automated Stereo Perception," Department of Computer Science, Stanford University, forthcoming Ph.D. thesis, 1982.

[Arvind 1982] Arvind, and Kim P. Gostelow, "The U-Interpreter," *Computer,* Vol. 15, No. 2, February 1982.

[Aslin 1980] Aslin, R.N., Dumais, S.T., Fox, R., Shea, S.L., "Stereopsis in human infants," *Science,* Vol. 207(4428), 323-324, 1980.

[Bajcsy 1973] Bajcsy, R., "Computer Identification of Visual Surface," *Computer Graphics and Image Processing,* vol. 2, 1973, 118-130.

[Baker 1980] Baker, H. Harlyn, "Edge Based Stereo Correlation," *Proc. ARPA Image Understanding Workshop,* University of Maryland, 168-175, April 1980.

[Baker 1981a] Baker, H. Harlyn and Thomas O. Binford, "Depth from Edge and Intensity Based Stereo," *Proceedings of the 7th International Joint Conference on Artificial Intelligence,* 631-636, Vancouver, British Columbia, August 1981.

[Baker 1981b] Baker, H. Harlyn, "Depth from Edge and Intensity Based Stereo," University of Illinois, Ph.D. thesis, September 1981.

[Baker 1982] Baker, H. Harlyn and Thomas O. Binford, "A System for Automated Stereo Mapping," *International Society for Photogrammetry and Remote Sensing, Commission II Symposium on Advances in Instrumentation for Processing and Analysis of Photogrammetric and Remotely Sensed Data,* Ottawa, Ontario, August 1982.

[Ballard 1976] Ballard, D.H. and J. Sklansky, "A ladder-structured decision tree for recognizing tumors in chest radiographs," *IEEE Trans. Computers*, vol. C-25, 5, May 1976, 503-513.

[Ballard 1978] Ballard, D., C. Brown, J. Feldman; "An approach to knowledge-directed image analysis," in **Computer Vision Systems**, A. Hanson, E. Riseman, eds, Academic Press, New York, 1978.

[Barlow 1967] Barlow, H.B., C. Blakemore, J.D. Pettigrew, "The neural mechanism of binocular depth discrimination," *Journal of Physiology*, 193, 327-342, 1967.

[Barnard 1979] Barnard, Stephen T. and William B. Thompson, "Disparity Analysis of Images," Computer Science Department, University of Minnesota, January 1979.

[Barnes 1968] Barnes, George H., Richard M. Brown, Maso Kato, David J. Kuck, Daniel L. Slotnick, and Richard A. Stokes, "The ILLIAC IV Computer," *IEEE Transactions on Computers*, Vol. C-17, No. 8, August 1968.

[Barrow 1971] Barrow, H.G., A.P. Ambler, R.M. Burstall, "Some techniques for recognizing Structures in Pictures," *Int. Conf on Frontiers in Pattern recognition*, Honolulu, Hawaii, Jan 1971.

[Batcher 1979] Batcher, K., "MPP-A Massively Parallel Processor," *Proc. International Conference on Parallel Processing*, August 1979.

[Batcher 1980] Batcher, Kenneth E., "Architecture of a Massively Parallel Processor," *Proc. 7th Annual Symposium on Computer Architecture*, May 1980.

[Beaudet 1978] Beaudet, P.R., "Rotationally invariant image operators," in*Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, (Kyoto, Japan, Nov. 7-10, 1978), 579-583.

[Bertram 1963] Bertram, S., "Automatic Map Compilation," *Photogrammetric Engineering*, January 1963.

[Bertram 1965] Bertram, S., "The Universal Automated Map Compilation Equipment," vol. 15, part 4, International Archives of Photogrammetry, 1965; *Photogrammetric Engineering*, 1965.

[Bertram 1969] Bertram, S., "The UNAMACE and the Automatic Photomapper," *Photogrammetric Engineering vol. 35*, no. 6, 569-576 June 1969.

[Binford 1970] Binford, T.O., "The TOPOLOGIST," Internal Report MIT-AI, 1970.

[Binford 1971] Binford, T.O., "Visual Perception by Computer," invited paper at the *IEEE Conference on Systems, Science and Cybernetics*, Miami, December 1971.

[Binford 1981] Binford, T.O., "Inferring Surfaces from Images," *Artificial Intelligence*, 17, 1981, 205-244.

[Bishop 1975] Bishop, Peter O., "Binocular Vision," in *Adler's Physiology of the Eye*, 558-614, Robert A. Moses, editor, The C. V. Mosby Company, St. Louis, 1975.

[Blachut 1976] Blachut, T.J., chairman, "Results of the International Orthophoto Experiment 1972-76," XIII Congress of the International Society Of Photogrammetry, Helsinki, 1976, publ. National Research Council Canada, NCR 15362, May 1976.

[Blake 1979] Blake, Randolph and Cormack, Robert, "Does contrast disparity alone generate stereopsis?," *Vision Research*, Vol. 19(8) 913-915, 1979.

[Blakemore 1970] Blakemore, Colin, "A New Kind of Stereoscopic Vision," *Vision Research*, vol. 10, 1970, 1181–1199.

[Blank 1981] Blank, Tom, Mark Stefik, and Williem vanCleemput, "A Parallel Bit Map Processor Architecture for DA Algorithms," *Proc. 18th Design Automation Conference*, 1981.

[Blum 1967] Blum, Harry, "A transformation for extracting new descriptors of shape," *Symposium on Models for Perception of Speech and Visual Form*, 362–380, ed. Weiant Whaten-Dunn, MIT Press, Cambridge, Mass., 1967.

[Bolles 1976] Bolles, R.C., "Verification Vision Within a Programmable Assembly System," AI Lab, Stanford University, Memo AIM-295, 1976.

[Brice 1970] Brice, C.R. and C.L. Fennema, "Scene Analysis Using Regions," Artificial Intelligence Group Technical Note 17, Stanford Research Institute, April 1970.

[Brode 1981] Brode, Brian, "Precompilation of Fortran Programs to Facilitate Array Processing," *Computer*, Vol. 14, No. 9, September 1981.

[Brooks 1981] Brooks, Rodney A., "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial Intelligence Journal*, vol. 16, 1981.

[Burr 1981] Burr, D.J., Bryan Ackland, Neil Weste, "A High Speed Array Computer for Dynamic Time Warping," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Atlanta, Ga., 471–474, March 1981.

[Burt 1980] Burt, Peter and Bela Julesz, "A Disparity Gradient Limit for Binocular Fusion," *Science*, vol. 208, no. 9, 615–617, May 1980.

[Butler 1978] Butler, Thomas and Westheimer, Gerald, "Interference with stereoscopic acuity: spatial, temporal, and disparity tuning," *Vision research*, Vol. 18(10), 1387-1392, 1978.

[Caelli 1978a] Caelli, T. and B. Julesz, "On Perceptual Analyzers Underlying Visual Texture Discrimination: Part I," *Biological Cybernetics*, vol. 28, 1978, 167–176.

[Caelli 1978b] Caelli, T., B. Julesz and E. Gilbert, "On Perceptual Analyzers Underlying Visual Texture Discrimination: Part II," *Biological Cybernetics*, vol. 29, no. 4, 1978, 201–214.

[Chapman 1975] Chapman, Dean R., Hans Mark, and Melvin W. Pirtle, "Computers vs. Wind Tunnels for Aerodynamic Flow Simulations," *Astronautics and Aeronautics*, April 1975.

[Charlesworth 1981] Charlesworth, Alan E., "An Approach to Scientific Array Processing: the Architectural Design of the AP-120B/FPS-164 Family," *Computer*, Vol. 14, No. 9, September 1981.

[Chen 1980] Chen, P.C. and T. Pavlidis, "Image Segmentation as an Estimation Problem," Computer Graphics and Image Processing, February 1980, vol. 12, no. 2, 153–172.

[Clarkson 1981] Clarkson, K.L., "A Procedure for Camera Calibration," *Proceedings DARPA Image Understanding Workshop*, 175–177, April 1981.

[Clocksin 1980] Clocksin, William F., "Perception of surface slant and edge labels from optical flow: a computational approach," *Perception*, vol. 9, 253-260, 1980.

[Cogan 1978] Cogan, Alexander, "Qualitative observations in visual science: "The Farnsworth Shelf:" Fusion at the site of the "ghosts,"" *Vision Research*, Vol. 18(6), 657-664, 1978.

[Comerford 1973] Comerford, James, "Stereopsis with chromatic contours," *Dissertation Abstracts International*, Vol. 34(2-B), 891, 1973.

[Conners 1980a] Conners, R.W. and C.A. Harlow, "A Theoretical Comparison of Texture Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 1980, 204-222.

[Conners 1980b] Conners, R.W. and C.A. Harlow, "Toward a Structural Textural Analyzer Based on Statistical Methods," *Computer Graphics and Image Processing*, March 1980, 224-256.

[Cooper 1980] Cooper, D.B., H. Elliott, F. Cohen, L. Reiss, and P. Symosek, "Stochastic Boundary Estimation and Object Recognition," *Computer Graphics and Image Processing*, vol.12, 1980, p. 326.

[Davis 1982] Davis, Alan L., and Robert M. Keller, "Data Flow Program Graphs," *Computer*, Vol. 15, No. 2, February 1982.

[Davis 1973] Davis, L., "A Survey of Edge Detection Techniques", TR-273, Univ of Md, Computer Science Center, 1973.

[Davis 1979a] Davis, L.S., "Computing the Spatial Structure of Cellular Textures," *Computer Graphics and Image Processing*, vol. 11, no. 2, October 1979, 111-122.

[Davis 1979b] Davis, L.S., S.A. Johns and J.K. Aggarwal, "Texture Analysis Using Generalized Co-occurrence Matrices," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 3, July 1979, 251-259.

[Davis 1980] Davis, L.S. and A. Mitiche, "Edge Detection in Textures," *Computer Graphics and Image Processing*, vol. 12, no. 1, Jan 1980, 25-39.

[DeBoor 1978] DeBoor, C., **A Practical Guide to Splines**, Springer, 1978 (Vol 27 in Applied Mathematical Sciences series).

[Degryse 1980] DeGryse, Donald G. and Dale J. Panton, "Syntactic Approach to Geometric Surface Shell Determination," *Soc. Photo-Optical Instrumentation Engineers*, vol. 238, 264-272, August 1980.

[Deguchi 1978] Deguchi, K. and I. Morishita, "Texture Characterization and Texture-based Image Partitioning Using Two-dimensional Linear Estimation Techniques," *IEEE Transactions on Electronic Computers*, vol. 27, no. 8, Aug. 1978, 739-745.

[de Weert 1979] de Weert, Charles, "Colour contours and stereopsis," *Vision Research*, Vol. 19(5), 555-564, 1979.

[Dineen 1955] Dineen, G.P., "Programming Pattern Recognition," *Proc. WJCC*, 94-100, March 1955.

[do Carmo 1976] do Carmo, M.P., **Differential Geometry of Curves and Surfaces**, Prentice-Hall, Englewood Cliffs, N.J., 1976.

[Dowman 1977] Dowman, I.J., "Developments in On Line Techniques of Photogrammetry and Digital Mapping," *Photogrammetric Record*, vol. 9, no. 49, 41–55, 1977.

[Dreschler 1981a] Dreschler, L. and H.-H. Nagel, "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," Report IfI-HH-M-90/81, Fachbereich Informatik, Universität Hamburg.

[Dreschler 1981b] Dreschler, L. and H.-H. Nagel, "Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene," *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (IJCAI-81), Aug 1981, Vancouver.

[Duda 1971] Duda, R.O. and P.E. Hart, "A generalized Hough transformation for detecting lines in pictures," SRI AI Group Tech Note 36, 1971.

[Duda 1972] Duda, R.O. and P.E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM* 15, no. 1, 1972, 11–15.

[Duda 1973] Duda, R.O. and P.E. Hart, **Pattern Classification and Scene Analysis**, Wiley, New York, 1973.

[Duff 1977] Duff, M. J. B., and D. M. Watson, "The Cellular Logic Array Image Processor", *The Computer Journal*, Vol. 20, No. 1, 1977.

[Duff 1978] Duff, M. J. B., "Review of the CLIP Image Processing System," *Proc. National Computer Conference*, 1978.

[Evans 1968] Evans, Thomas G., "A heuristic Program to Solve Geometric Analogy Problems," **Semantic Information Processing**, ed. Marvin Minsky, MIT Press, Cambridge, Mass., 1968.

[Eversole 1980] Eversole, William L., and Dale J. Mayer, "Investigation of VLSI Technologies for Image Processing," *Proc. Image Understanding Workshop*, April 1980.

[Faugeras 1980] Faugeras, O., and K. Price, "Semantic Description of Aerial Images Using Stochastic Labelling," *Proc ARPA Image Understanding Workshop*, 89, Univ of Md, April 1980.

[Faugeras 1980] Faugeras, O.D. and W.K. Pratt, "Decorrelation Methods of Texture Feature Extraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, July 1980, 323–332.

[Fennema 1970] Fennema, C.L. and C.R. Brice, "Scene analysis of pictures using regions", *Artificial Intelligence Journal* 1, 1970, 205–226.

[Fennema 1979] Fennema, Claude L. and William B. Thompson, "Velocity Determination in Scenes Containing Several Moving Objects," *Computer Graphics and Image Processing*, 9, 1979.

[Foley 1975] Foley, J.M., Applebaum, T.H., Richards, W.A., "Stereopsis with large disparities: Discrimination and depth magnitude," *Vision Research* 15, 417-421, 1975.

[Forney 1973] Forney, G. David Jr., "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, no. 3, 268-278, March 1973.

[Fox 1977] Fox, Robert, Lehmkuhle, Stephen and Bush, Robert, "Stereopsis in the falcon," *Science*, Vol. 197(4298), 79-81, 1977.

[Friedman 1978] Friedman, Rhonda, Kaye, Martin, Richards, Whitman, "Effect of vertical disparity upon stereoscopic depth," *Vision Research*, Vol. 18(3), 351-352, 1978.

[Friedman 1980] Friedman, S.J., editor, Manual of Photogrammetry, American Society of *Photogrammetry*, C. C. Slama, editor-in-chief, 1980.

[Frisby 1975] Frisby, John and Julesz, Bela, "The effect of orientation differences on stereopsis as a function of line length," *Perception*, Vol. 4(2), 179-186, 1975.

[Frisby 1976] Frisby, John and Julesz, Bela, "The effect of length differences between corresponding lines on stereopsis from single and multi-line stimuli," *Vision Research*, Vol. 16(1), 83-87, 1976.

[Frisby 1977a] Frisby, John P. and John E.W. Mayhew, "Global Processes in Stereopsis: Some Comments on Ramachandran and Nelson(1976)," *Perception*, vol. 6, 195–206, 1977.

[Frisby 1977b] Frisby, John and Mayhew, John, "Global processes in stereopsis: some comments on Ramachandran and Nelson," *Perception*, Vol. 6(2), 195-206, 1977.

[Frisby 1978] Frisby, John and Mayhew, John, "Contrast sensitivity function for stereopsis," *Perception*, Vol. 7(4), 423-429, 1978.

[Frisby 1979] Frisby, John and Mayhew, John, "Depth inversion in random-dot stereogram," *Perception*, Vol. 8(4), 397-399, 1979.

[Fu 1978] Fu, K. S., "Special Computer Architectures for Pattern Recognition and Image Processing — an Overview," *Proc. National Computer Conference*, 1978.

[Fuller 1976] Fuller, Samuel H., "Price/Performance Comparison of C.mmp and the PDP-10," *Proc. 3rd Annual Symposium on Computer Architecture*, January 1976.

[Gagalowicz 1980] Gagalowicz, A., "Visual Discrimination of Stochastic Texture Fields based upon their second Order Statistics," *Proceedings of Fifth International Conference on Pattern Recognition*, Miami Beach, Fl., December 1980, 786–788.

[Gajski 1982] Gajski, D. D., D. A. Padua, D. J. Kuck, and R. H. Kuhn, "A Second Opinion on Data Flow Machines and Languages," *Computer*, Vol. 15, No. 2, February 1982.

[Garber 1981] Garber, D., "Computational Models for Texture Analysis and Texture Synthesis," University of Southern California, USCIPI Report 1000, May 1981, (Ph.D. Thesis).

[Garvey 1976] Garvey, T.D., "Perceptual Strategies for Purposive Vision," SRI AI Center Tech Note 117, 1976.

[Gennery 1980] Gennery, Donald B., "Modelling the Environment of an Exploring Vehicle by Means of Stereo Vision," Ph.D. thesis, Stanford Artificial Intelligence Laboratory, AIM-339, June 1980.

[Gibson 1950] Gibson, James J., *"The Perception of the Visual World,"* The Riverside Press, Houghton Mifflin Co., 1950.

[Gimel'farb 1972] Gimel'farb, G.L., V.B. Marchenko, and V.I. Rybak, "An Algorithm for Automatic Identification of Identical Sections on Stereopair Photographs," *Kybernetica* (translations) no. 2, 311–322, March–April 1972.

[Gregory 1977] Gregory, Richard, "Vision with isoluminant colour contrast: 1. a projection technique and observations," *Perception*, 6, 113, 1977.

[Griffith 1973] Griffith, A.K., "Mathematical Models for Automatic Line Detection," *Journal of the ACM*, vol. 20, no. 1, January 1973, p. 62.

[Grimson 1980] Grimson, W.E.L., "Computing Shape Using a Theory of Human Stereo Vision," Department of Mathematics, MIT, June 1980.

[Guibas] Guibas, Leo, H. T. Kung, and Thompson, "Direct VLSI Implementation of Combinatorial Algorithms," extended abstract.

[Habibi 1972] Habibi, A., "Two Dimensional Bayesian Estimate of Images", *Proceedings of the IEEE*, vol. 60, no. 7, 1972, 878–883.

[Hallert 1960] Hallert, Bertil, *"Photogrammetry, Basic Principles and General Survey,"* McGraw-Hill Book Company Inc., 1960.

[Halmos 1957] Halmos, P.R., **Introduction to Hilbert Space and the Theory of Spectral Multiplicity**, Chelsea, 1957.

[Halmos 1963] Halmos, P.R., "What Does the Spectral Theorem Say?," *The American Mathematical Monthly*, March 1963, 241–247.

[Hamsher 1978] Hamsher, K., "Stereopsis and the perception of anomalous contours," *Neuropsychologia*, Vol. 16(4), 453-459, 1978.

[Hannah 1974] Hannah, Marsha Jo, "Computer Matching of Areas in Stereo Images," Ph.D. thesis, Stanford Artificial Intelligence Laboratory, AIM–239, July 1974.

[Haralick 1973] Haralick, R.M., K. Shanmugam and I. Dinstein, "Texture Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, Nov. 73, 610–621.

[Haralick 1979] Haralick, R.M., "Statistical and Structural Approaches to Texture," *Proceedings of IEEE*, vol. 67, no. 5, May 1979, 786–804.

[Haralick 1980] Haralick, R. M., "Edge and Region Analysis for Digital Image Data," *Computer Graphics and Image Processing*, vol. 12, no. 1, January 1980, 60–73.

[Haralick 1981] Haralick, R. M., "The Digital Edge," *Proc. IEEE Conf. Pattern Recognition and Image Processing*, August 1981, 285–291.

[Harris 1973] Harris, J.P. and Gregory, R.L., "Fusion and rivalry of illusory contours," *Perception*, Vol. 2(2), 235-247, 1973.

[Hawkins 1970] Hawkins, J.K., "Texture Properties for Pattern Recognition," in **Picture Processing and Psychopictorics**, B.S. Lipkin and A. Rosenfeld, (Editors), Academic Press, New York, 1970, 347–370.

[Helava 1957] Helava, U.V., International Photogrammetric Conference on Aerial Triangluation, Ottawa, August 1957.

[Henderson 1979a] Henderson, Robert L., Walter J. Miller, C.B. Grosch, "Automatic Stereo Recognition of Man-Made Targets," *Society of Photo-Optical Instrumentation Engineers*, August 1979.

[Henderson 1979b] Henderson, R.L., "Geometric Reference Preparation Interim Report Two: The Broken Segment Matcher," RADC-TR-79-80, April 1979.

[Herman 1974] Herman, John, Tauber, Edward, Roffwarg, Howard, "Monocular occlusion impairs stereoscopic acuity, but total visual deprivation does not," *Perception and Psychophysics*, Vol. 16(2), 225-228, 1974.

[Herskovits 1970] Herskovits, A. and T.O. Binford, "On Boundary Detection," MIT Project MAC, Artificial Intelligence Memo 183, July 1970.

[Hewitt 1968] Hewitt, C., "Planner," MIT AI Memo 168, 1968.

[Hobrough 1978] Hobrough, G.L. and T.B. Horbrough, "Image On-line correlation," *Bildmessung und Liftbildwessen*, vol. 46, no. 3, 79-86, 1978.

[Hogben 1976] Hogben, J.H., Julesz, Bela and Ross, John, "Short-term memory for symmetry," *Vision Research*, Vol. 16(8), 861-866, 1976.

[Horn 1972] Horn, B.K.P., "The Binford-Horn Edge Finder," MIT AI Memo 285, 1972, revised December 1973.

[Horn 1973] Horn, B.K.P., "On Lightness," MIT-AI Memo 295, 1973.

[Hough 1962] Hough, P.V.C., "Method and Means for recognizing complex patterns," U.S.Patent 3,069,654, December 18, 1962.

[Hsu 1978] Hsu, S., J.L. Mundy, P.R. Beaudet, "Web Representation of Image Data," *Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, (Kyoto, Japan, Nov. 7-10, 1978), 579-583.

[Hubel 1962] Hubel, D.H. and T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, 160, 385-398, 1962.

[Hubel 1970] Hubel, D.H. and T.N. Wiesel, "Cells sensitive to binocular depth in area 18 of the macaque monkey cortex," *Nature*, 225, 41-42, 1970.

[Hueckel 1969] Hueckel, M.H., "An Operator which Locates Edges in Digital Pictures," Stanford Computer Science Dept. Memo AIM-105, Oct. 1969.

[Hueckel 1971] Hueckel, M.H., "An Operator which Locates Edges in Digital Pictures," *JACM*, vol. 18, no. 1, January 1971, 113-125. Erratum in 21, 1974,350.

[Hueckel 1973] Hueckel, M., "A Local Visual Operator which Recognizes Edges and Lines," *J.Assoc Computing Machinery*, 20, 634 (1973).

[Iversen 1981] Iversen, Wesley R., "Total Immersion Cools Supercomputer Logic," *Electronics*, Vol. 54, No. 24, November 1981.

[Jacobus 1980] Jacobus, Charles J., Robert T. Chien, John Michael Selander, "Motion Detection and Analysis of Matching Graphs of Intermediate-Level Primitives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, Nov 1980.

[Julesz 1962] Julesz, B., "Visual Pattern Discrimination," *IRE Transactions on Information Theory*, vol. 8, February 1962, 84–92.

[Julesz 1964] Julesz, Bela, "Binocular depth perception without familiarity cues," *Science*, 145(3630) 1964.

[Julesz 1971] Julesz, Bela, *"Foundations of Cyclopean Perception,"* Chicago, University of Chicago Press, 1971.

[Julesz 1974] Julesz, Bela, "Cooperative phenomena in binocular depth perception," *American Scientist*, Vol. 62(1), 32-43, 1974.

[Julesz 1975] Julesz, B., "Experiments in the Visual Perception of Texture," *Scientific American*, Apr. 1975, 34–43.

[Julesz 1975a] Julesz, Bela and Miller, Joan, "Independent spatial-frequency-tuned channels in binocular fusion and rivalry," *Perception*, Vol. 4(2), 125-143, 1975.

[Julesz 1975b] Julesz, Bela, "Two-dimensional spatial-frequency-tuned channels in visual perception," Proc. of Intern. Conf. on Signal Analysis and Pattern Recognition in *Biomedical Engineering*, 177-197, 1975.

[Julesz 1976] Julesz, Bela, "Global Stereopsis: Cooperative Phenomena in Stereoscopic Depth Perception," in *Handbook of Sensory Physiology VIII*, R. Held, H. Leibovitz and H-L. Teuber, editors, Springer, Berlin, 1976.

[Julesz 1978] Julesz, B., E.N.Gilbert and J.D. Victor, "Visual Discrimination of Textures with Identical Third-Order Statistics," *Biological Cybernetics*, vol. 31, no. 3, 1978, 137–140.

[Julesz 1978] Julesz, Bela, "Binocular utilization of monocular cues that are undetectable monocularly," *Perception*, Vol. 7(3), 315-322, 1978.

[Kanade 1978] Kanade, T., "Region Segmentation: Signal vs. Semantics," *Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, (Kyoto, Japan, Nov. 7–10, 1978), 579–583.

[Kanade 1981] Kanade, T., and R. Reddy, "Image Understanding at CMU," *Proc IU Workshop*, April 1981.

[Kanade 1981] Kanade, T., "Recovery of the Three-Dimensional Shape of an Object from a Single View," *Artificial Intelligence*, vol. 17, 409, 1981.

[Karplus 1981] Karplus, Walter J., and Danny Cohen, "Architectural and Software Issues in the Design and Application of Peripheral Array Processors," *Computer*, Vol. 14, No. 9, September 1981.

[Kaufman 1976] Kaufman, Lloyd, "On stereopsis with double images," *Psychologia: An International Journal of Psychology in the Orient*, Vol. 19(4), 224-233, 1976.

[Kelley 1970] Kelley, Michael D., "Visual Recognition of People by Computer," *Stanford Artificial Intelligence Laboratory, AIM-130, CS-168*, Ph.D. thesis, 1970.

[Kelly 1977] Kelly, R.E., P.R.H. McConnell, and S.J. Mildenberger, "The Gestalt Photomapping System," *Journal of Photogrammetric Engineering and Remote Sensing*, vol. 43, 1407, 1977.

[Kidd 1979] Kidd, A.L., Frisby, J.P., Mayhew, J.E.W., "Texture contours can facilitate stereopsis by initiating vergence eye movements," *Nature*, Vol. 280(5725), 829-832, 1979.

[Kirsch 1971] Kirsch, R.A., "Computer Determination of the Constituent Structure of Biological Images," *Computers and Biomedical Research*, vol. 4, no. 3, 1971, 315–328.

[Klein 1977] Klein, Raymond, "Stereopsis and the representation of space," *Perception*, Vol. 6(3), 327-332, 1977.

[Koenderink 1976] Koenderink, J.J.and Van Doorn, A.J., "Geometry of binocular vision and a model for stereopsis," *Biol. Cybern.*, (Germany), Vol. 21(1), 29-35, 1976.

[Kopetzky 1980] Kopetzky, Daniel J., "An Array Simulator Generator," UILU-ENG 80-1737, Ph.D. thesis, University of Illinois at Urbana-Champaign, September 1980.

[Krulicoski 1972] Krulicoski, S.J. and R.B. Forest, "Coherent Optical Terrain Relief Determination using a Matched Filter," *Bendix Technical Journal*, vol. 5, no. 1, Spring 1972.

[Kuck 1968] Kuck, David J., "Illiac IV Software and Application Programming," *IEEE Trans. on Computers*, Vol. C-17, No. 8, August 1968.

[Kuck 1976] Kuck, David J., "Parallel Processing of Ordinary Programs," *Advances in Computers*, Vol. 15, 1976.

[Kulikowski 1978] Kulikowski, J., "Limit of single vision in stereopsis depends on contoursharpness," *Nature*, Vol. 275(5676), 126-127, 1978.

[Kung 1980] Kung, H. T., "The Structure of Parallel Algorithms," *Advances in Computers*, Vol. 19, 1980.

[Kushner 1980] Kushner, Todd, Angela Y. Wu, and Azriel Rosenfeld, "Image Processing on ZMOB," TR-987, University of Maryland Computer Science Center, December 1980.

[Lampson 1981] Lampson, Butler W., and Kenneth A. Pier, "A Processor for a High-Performance Personal Computer," appeared in "The Dorado: A High-Performance Personal Computer, Three Papers," CSL-81-1, Xerox PARC, January 1981.

[Land 1977] Land, E.H.; "The Retinex Theory of Color Vision," *Scientific American*, 1977.

[Landau 1961] Landau, H.J. and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — II," *Bell Syst. Tech. J.*, 40, January 1961, 65–84.

[Landau 1962] Landau, H.J. and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — III: The Dimension of the Space of Essentially Time- and Band-Limited Signals," *Bell Syst. Tech. J.*, 41, July 1962, 1295–1336.

[Laws 1980] Laws, K.I., "Textured Image Segmentation," University of Southern California Report USCIPI 940 (Ph.D. thesis), Jan. 1980.

[Lawton 1981] awton, Daryl T., "Optic flow field structure and processing image motion," *Seventh Int. Joint Conf. on Artificial Intelligence*, Vancouver, B.C., 700–703, August 1981.

[Lema 1977] Lema, Sandra and Blake, Randolph, "Binocular summation in normal and stereoblind humans," *Vision Research*, Vol. 17(6), 691-695, 1977.

[Levine 1973] Levine, Martin D., Douglas A. O'Handley, Gary M. Yagi, "Computer Determination of Depth Maps," *Computer Graphics and Image Processing*, 2, 131–150, 1973.

[Levine 1978] Levine, M., "A knowledge-based computer vision system," in **Computer Vision Systems**, A. Hanson, E. Riseman, eds, Academic Press, New York, 1978.

[Levine 1982] Levine, Ronald D., "Supercomputers," *Scientific American* Vol. 246, No. 1, January 1982.

[Levinson 1979] Levinson, Eugene and Blake, Randolph, "Stereopsis by harmonic analysis," *Vision Research*, Vol. 19(1), 73-78, 1979.

[Liebes 1977] Liebes Jr., S. and A.A. Schwartz, "Viking 1975 Mars Lander Interactive Computerized Video Stereophotogrammetry," *Journal of Geophysics Research*, 82, no. 28, 4421, Sept. 30, 1977.

[Liebes 1981] Liebes Jr., S., "Geometric Constraints for Interpreting Images of Common Structural Elements: Orthogonal Trihedral Vertices," *Proceedings of the ARPA Image Understanding Workshop*, 1981.

[Lowe 1981] Lowe, D.G., and T.O. Binford, "The Interpretation of Geometric Structure from Image Boundaries," *Proc. ARPA Image Understanding Workshop*, 39–46, April 1981.

[Lowry 1981] Lowry, Michael R. and Allan Miller, "A General Purpose VLSI Chip for Computer Vision with Fault-Tolerant Hardware," *Proc. ARPA Image Understanding Workshop*, Washington D. C., April 1981, 184–187.

[Lu 1971] Lu, C. and Fender, D., "Interaction of color and luminance in stereoscopic vision," 1971 Annual Meeting of the Optical Society of America.

[Lucas 1981] Lucas, Bruce D., and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Seventh Int. Joint Conf. on Artificial Intelligence*, Vancouver, British Columbia, 674–679, August 1981.

[Luria 1975] Luria, S. and Kinney, J., "Vision in the water without a facemask," *Aviation, Space and Environmental Medicine*, Vol. 46, 1975.

[Machuca 1981] Machuca, R. and A.L. Gilbert, "Finding edges in noisy scenes," *Pattern Analysis and Machine Intelligence*, PAMI-3, no. 1, January 1981, p. 303.

[Macrenhas 1978] Macrenhas, N.D.A. and L.O.C.Prado, "Edge detection in images: a hypothesis testing approach," in *Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR-78)*, Kyoto, Japan, Nov.7-10, 1978.

[Maleson 1977] Maleson, J.T., C.M. Brown and J.A. Feldman, "Understanding Natural Texture," *Proceedings of ARPA Image Understanding Workshop*, Palo Alto, Ca., October 1977, 19-27.

[Manuel 1981a] Manuel, Tom, "Japanese Map Computer Domination," *Electronics*, Vol. 54, No. 23, November 1981.

[Manuel 1981b] Manuel, Tom, "West Wary of Japan's Computer Plan," *Electronics*, Vol. 54, No. 25, December 1981.

[Marks 1980] Marks, Philip, "Low-Level Vision Using an Array Processor," *Computer Graphics and Image Processing, no. 14, 1980, 281–292.*

[Maron 1981] Maron, Neil, and Thomas A. Brengle, "Integrating an Array Processor into a Scientific Computing System," *Computer*, Vol. 14, No. 9, September 1981.

[Marr 1974] Marr, David, "On The Purpose of Low-Level Vision," MIT Artificial Intelligence Memo no. 324, December 1974.

[Marr 1976a] Marr, D., "Early Processing of Visual Information," *Philosophical Transactions of the Royal Society*, London,Series B, 275, p 483–524, 1976.

[Marr 1976b] Marr, D. and T. Poggio, " Cooperative Computation of Stereo Disparity," *Science*, vol. 194, October 1976, 283–287.

[Marr 1977] Marr, D. and T. Poggio, "A Theory of Human Stereo Vision," MIT Artificial Intelligence Memo no. 451, November 1977.

[Marr 1978] Marr, D., Palm, G., Poggio, T., "Analysis of a cooperative stereo algorithm," *Biol. Cybern.*, Vol. 28(4), 223-229, 1978.

[Marr 191979] Marr, D. and E. Hildreth, "Theory of Edge Detection," AI Memo 518, MIT AI Lab, April 1979. Also Proc.R.Soc.Lond.B., 1980, 207, 187–217.

[Marr 79b] Marr, D., T. Poggio, S. Ullman, "Bandpass Channels, Zero-crossings, and Early Visual Information Processing," *Journal of the Optical Society of America*, 1979.

[Martelli 1972] Martelli, A., "Edge Detection using Heuristic Search Methods," Dept of EE and Computer Science, NYU, University Heights, Bronx, NY, 10453. Also *Computer Graphics and Image Processing*, 1, 1972, 169–182.

[Martelli 1973] Martelli, A., "An Application of Heuristic Search Methods to Edge and Contour Detection," Instituto di Elaborazione della Informazione del Consiglio Nazionale delle Richerche, Pisa, 1973. Also *Comm. ACM* 19, 1976, 73–83.

[Mayhew 1977] Mayhew, John, Frisby, John, and Gale, Peter, "Computations of stereo disparity from rivalrous texture stereograms," *Perception*, Vol. 6(2), 207-208, 1977.

[Mayhew 1978a] Mayhew, John and Frisby, John, "Stereopsis masking in humans is not orientationally tuned," *Perception*, Vol. 7(4), 431-436, 1978.

[Mayhew 1978b] Mayhew, John and Frisby, John, "Contrast summation effects and stereopsis," *Perception*, Vol. 7(5), 537-550, 1978.

[Mayhew 1979] Mayhew, John and Frisby, John, "Surfaces with steep variations in depth pose difficulties for orientationally tuned disparity filters," *Perception*, Vol. 8(6), 691-698, 1979,.

[Mayhew 1980] Mayhew, John, "The computation of binocular edges," *Perception*, Vol. 9(1), 69-68, 1980,.

[Mayhew 1981] Mayhew, John E.W. and John P. Frisby, "Computational and Psychological Studies Towards a Theory of Human Stereopsis," *Artificial Intelligence Journal*, vol. 16, 1981.

[McCormick 1974] McCormick, B.H. and S.N. Jayaramamurthy, "Time Series Model for Texture Synthesis," *International Journal of Computer and Information Science*, vol. 3, no. 4, 1974, 329–343.

[McCormick 1975] McCormick, B.H. and S.N. Jayaramamurthy, "A Decision Theory Method for the Analysis of Texture," *International Journal of Computer and Information Science*, vol. 4, no. 1, 1975, 1–37.

[Meiri 1981] Meiri, A. Zvi, "On Monocular Perception of 3-D Moving Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, Nov. 1980.

[Mitchell 1970] Mitchell, D.E. and Blakemore, C., "Binocular depth perception and the corpus callosum," *Vision Research*, Vol. 10(1), 49-54, 1970.

[Mitchell 1973] Mitchell, Donald and Baker, Andrew, "Stereoscopic aftereffects: Evidence for disparity-specific neurones in the human visual system," *Vision Research*, Vol. 13(12), 2273-2288, 1973.

[Miyamoto 1975] Miyamoto, E., and T.O. Binford, "Display Generated by a Generalized Cone Display," *Proc. Conf. on Computer Graphics, Pattern recognition, and Data Structures*, May, 1975.

[Modestino 1980] Modestino, J.W., R.W. Fries and A.L. Vickers, "Stochastic Image Models Generated by Random Tesselations of the Plane," *Computer Graphics and Image Processing*, vol. 12, 1980, 74–98.

[Montanari 1970] Montanari, U., "On the Optimal Detection of Curves in Noisy Pictures," Artificial Intelligence Laboratory, Stanford University, Memo AIM-115, 1970.

[Montanari 1971] Montanari, U., "On the Optimal Detection of Curves in Noisy Pictures," *Comm. ACM* 14, May 1971, 335–345.

[Moore 1979] Moore, Roger K., "A Dynamic Programming Algorithm for the Distance Between Two Finite Areas," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 1, January 1979, 86–88.

[Moravec 1977] Moravec, H.P., "Towards Automatic Visual Obstacle Avoidance," *Proc 5th IJCAI*, 1977.

[Moravec 1980] Moravec, Hans P., "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover," Stanford Artificial Intelligence Laboratory, AIM–340, Ph.D. thesis, September 1980.

[Mori 1973] Ken-Ichi Mori, Masatsugu Kidode, Haruo Asada, "An Iterative Prediction and Correction Method for Automatic Stereocomparison," *Computer Graphics and Image Processing*, 2, 393-401, 1973.

[Morse 1953] Morse, P.M. and H. Feshbach, **Methods of Theoretical Physics, Part II**, McGraw-Hill, 1953.

[Movshon 1972] Movshon, J.A., Chambers, B.E., Blakemore, C., "Interocular transfer in normal humans and those who lack stereopsis," *Perception*, Vol. 1(4), 483-490, 1972.

[Nagao 1978] Nagao, M., T. Matsuyama, Y. Ikeda; "Region Extraction and Shape Analysis of Aerial Photographs," *Proc 4ICPR*, p 620, 1978.

[Nagao 1980] Nagao, M. and T. Matsuyama, **A structural Analysis of Complex Aerial Photographs**, Plenum Press, New York, 1980.

[Nagel 1981] Nagel, Hans-Helmut, and Bernd Neumann, "On 3D Reconstruction from Two Perspective Views," *Seventh Int. Joint Conf. on Artificial Intelligence*, Vancouver, B.C., 661-663, August 1981.

[Nelson 1975] Nelson, J.I., "Globality and stereoscopic fusion in binocular vision," *J. of Theoretical Biology*, Vol. 49(1), 1-88, 1975.

[Nelson 1977] Nelson, J.I., H. Kato, and P.O.Bishop, "Discrimination of orientation and position disparities by binocularly activated neurons in cat striate cortex," *Journal of Neurophysiology*, 40(2):260-283 1977.

[Neumann 1980] Neumann, Bernd, "Exploiting Image Formation Knowledge for Motion Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, November 1980.

[Nevatia 1974] Nevatia, Ramakant, "Structured Descriptions of Complex Curved Objects for Recognition and Visual Memory," Ph.D. Dissertation, Dept. of Computer Science, Stanford University, STAN-CS-74-464, October 1974.

[Nevatia 1976] Nevatia, Ramakant, "Depth Measurement by Motion Stereo," *Computer Graphics and Image Processing*, 5, 1976.

[Nevatia 1977] Nevatia, R., and T.O. Binford, "Description and recognition of Curved Objects," *Artificial Intelligence Journal*, 1977.

[Nevatia 1977] Nevatia, R., "A Color Edge Detector and Its Use in Scene Segmentation," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-7, no 11, November 1977, 820-826.

[Nevatia 1978] Nevatia, R. and K.R. Babu, "Linear Feature Extraction," *Proc. ARPA Image Understanding Workshop*, Pittsburgh, November 1978, 73-78.

[Nikara 1968] Nikara, T., P.O. Bishop, J. Pettigrew, "Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex," *Exp. Brain Research*, 6, 353-372, 1968.

[Nishihara 1981] Nishihara, H.K., and N.G. Larson, "Towards a Real Time Implementation of the Marr and Poggio Stereo Matcher," *Proceedings of the ARPA Image Understanding Workshop*, 114-120, May 1981.

[O'Gorman 1976] O'Gorman, F., "Edge Detection using Walsh Functions," *Proc AISB*, p 195, July 1976. Also: *Artificial Intelligence* 10, 1978, 215-233.

[Ohlander 1975] Ohlander, R.B., "Analysis of Natural Scenes," Dept of Computer Science, Carnegie-Mellon Univ, April 1975. (PhD thesis)

[Ohta 1980] Ohta, Y., "A region-oriented image-analysis system by computer," Thesis, Dept of Information Science, Kyoto University, 1980.

[Panton 1978] Panton, Dale J., "A Flexible Approach to Digital Stereo Mapping," *Photogrammetric Engineering and Remote Sensing*, vol. 44, no. 12, 1499-1512, December 1978.

[Panton 1981] Panton,D.L., C.B. Grosch, D.G. DeGryse, J. Ozils, A.E. LaBonte, S.B. Kaufmann, L. Kirvida, "Geometric Reference Studies," RADC-TR-81-182, Final Technical Report, July 1981.

[Parma 1980] Parma, C.C., A.M. Hanson, E.M. Riseman; "Experiments in Schema-Driven Interpretation of a Natural Scene," Univ of Mass COINS Tech Rept 80-10, 1980.

[Pastore 1972] Pastore, Nicholas, "Sebastien Le Clerc on retinal disparity," *J. of the History of the Behavioral Sciences,* Vol. 8(3), 336-339, 1972.

[Pavlidis 1972] Pavlidis, T., "Segmentation of Pictures and Maps through Functional Approximation," *Computer Graphics and Image Processing,* vol. 1, 1972, 360-372.

[Pavlidis 1977] Pavlidis,T., **Structural Pattern Recognition,** Springer-Verlag, 1977.

[Pickett 1970] Pickett, R.M., "Visual Analysis of Texture in the Detection and Recognition of Objects," in *Picture Processing and Psychopictorics,* B.S. Lipkin and A. Rosenfeld, (Editors), Academic Press, New York, 1970, 289-308.

[Prazdny 1981a] Prazdny, K., "Relative Depth and Local Surface Orientation from Image Motions," *Proceedings of the ARPA Image Understanding Workshop,* 47-60, May 1981.

[Prazdny 1981b] Prazdny, K., "A simple method for recovering relative depth map in the case of a translating sensor," *Seventh Int. Joint Conf. on Artificial Intelligence,* Vancouver, B.C., 698-699, August 1981.

[Prazdny 1981c] Prazdny, K., "Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinear Moving Observer," *Computer Graphics and Image Processing,* 17, 238-248, 1981.

[Prewitt 1970] Prewitt, J.M.S., "Object Enhancement and Extraction," in **Picture Processing and Psychopictorics,** B.S.Lipkin and A.Rosenfeld,Eds., Academic Press, New York, 1970.

[Purks 1977] Purks, S.R. and W. Richards, "Visual Texture Discrimination Using Random Dot Patterns," *Journal of Optical Society of America,* vol. 67, June 1977, 765-771.

[Quam 1971] Quam, Lynn H., "Computer Comparison of Pictures," *Stanford Artificial Intelligence Laboratory, AIM-144,* Ph.D. thesis, 1971.

[Ramachandran 1973] Ramachandran, V., Rao, B. Madhusudhan, Sriram, S., Vidyasagar, T.R., "The role of colour perception and pattern recognition in stereopsis," *Vision Research,* Vol. 13(2), 505-509, 1973.

[Ramachandran 1976] Ramachandran, V.S. and Nelson, J.I., "Global grouping overrides point-to-point disparities," *Perception* 5, 125-128, 1976.

[Rashid 1981] Rashid, Richard F., "Towards a System for the Interpretation of Moving Light Displays," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 2, no. 6, November 1980.

[Regan 1973] Regan, D. and Beverley, I., "Electrophysiological evidence for existence of neurones sensitive to direction of depth movement," *Nature,* Vol. 246(5434), 504-506, 1973.

[Regan 1974] Regan, D. and Beverley, I., "Visual Sensitivity to Disparity Pulses: Evidence for Directional Selectivity," *Vision Research*, Vol. 14, 357-361, 1974.

[Richards 1970] Richards, Whitman, "Stereopsis and Stereoblindness," *Exp. Brain Research*, vol. 10, 1970, 380–388.

[Richards 1974a] Richards, Whitman and Kaye, Martin, "Local versus global stereopsis: Two mechanisms?," *Vision Research*, Vol. 14(2), 1345-1347, 1974.

[Richards 1974b] Richards, Whitman and Foley, John, "Effect of luminance and contrast on processing large disparities," *J. of the Optical Soc. of America*, Vol. 64(12), 1703-1705, 1974.

[Richards 1977] Richards, Whitman, "Stereopsis with and without monocular contours," *Vision Research*, Vol. 17(8), 967-969, 1977.

[Richards 1978] Richards, W., "Mechanisms for stereopsis" *Frontiers in Visual Science*, 387-395, 1978.

[Roach 1981] Roach, John W., and J.K. Aggarwal, "Determining the Movement of Objects from a Sequence of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, November 1980.

[Roberts 1963] Roberts, L.G., "Machine Perception of Three-Dimensional Solids," in **Optical and Electro-Optical Information Processing, Optical and Electro-Optical Information Processing**, J.T.Tippett et al., Eds., MIT Press, Cambridge, Mass., 1965, 159–197. Also Technical Report no. 315, Lincoln Laboratory, MIT (May 1963).

[Rogers 1975] Rogers, Brian and Anstis, Stuart, "Reversed depth from positive and negative stereograms," *Perception*, Vol. 4(2), 193-201, 1975.

[Rosenfeld 1975] Rosenfeld, A., R.A. Hummel, S.W. Zucker, "Scene Labelling by Relaxation Operations," Computer Science Center, Univ of Md, TR-379, May 1975. Also *IEEE Trans. Syst. Man Cybern.*, SMC-6, no. 6, June 1976, 420-433.

[Rosenfeld 1976] Rosenfeld, A. and A.C. Kak, **Digital Picture Processing**, Academic Press, New York, 1976.

[Ross 1975] Ross, John and Hogben, J., "The Pulfrich effect and short-term memory in stereopsis," *Vision Research*, Vol. 15(11), 1289-1290, 1975.

[Rubin 1978] Rubin, S., "The ARGOS Image Understanding System," *Proc ARPA IU Workshop*, Nov 1978; also "The ARGOS Image Understanding System," Ph.D. Thesis, Carnegie-Mellon University, 1978.

[Rubin 1980] Rubin, Steven M., "Natural Scene Recognition Using Locus Search," *Computer Graphics and Image Processing*, vol. 13, no. 4, 298–333, August 1980.

[Russell 1979] Russell, P.W., "Chromatic input to stereopsis," *Vision Research*, 19(7), 831-834, 1979.

[Rutkowski 1978] Rutkowski, W.S. and A. Rosenfeld, "A Comparison of Corner-Detection Techniques for Chain-Coded Curves," University of Maryland Technical Report TR-623, Jan. 1978.

[Ryan 1979] Ryan, T.W., R.T. Gray, and B.R. Hunt, "Prediction of Correlation Errors in Stereo-Pair Images," SIE/DIAL–79–002.

[Ryan 1980] Ryan, Thomas W., and B.R. Hunt, "The Prediction of Accuracy in Digital Cross-Correlation of Stereo-Pair Images," *Soc. Photo-Optical Instr. Engineers*, vol. 219, Electro-Optical Technology for Autonomous Vehicles, 1980.

[S-1 1979] The S-1 Project, "Fiscal Year 1979 Annual Report," UCID-18619, Lawrence Livermore National Laboratory, September 1979.

[Santalo 1976] Santalo Sors, L.A., "Integral Geometry and Geometric Probability," Addison-Wesley, 1976 (Vol 1 in Encyclopedia of Mathematics and its Applications).

[Saye 1975] Saye, Ann and John P. Frisby, "The Role of Monocularly Conspicuous Features in Facilitating Stereopsis from Random-Dot Stereograms," *Perception*, vol. 4(2), 159–171, 1975.

[Scarano 1976] Scarano, Frank A., "A Digital Elevation Data Collection System," *Photogrammetric Engineering and Remote Sensing*, vol. 42, no. 4, 489, April 1976.

[Schacter 1979] Schachter, B. and N. Ahuja, "Random Pattern Generation Process," *Computer Graphics and Image Processing*, vol. 10, 1979, 95–114.

[Schatz 1977] Schatz, B.R., "The Computation of Immediate Texture Discrimination," MIT AI Memo 426, August 1977.

[Schumer 1979] Schumer, Robert, "Independent stereoscopic channels for different extents of spatial pooling," *Vision Research*, Vol. 19(12), 1303-1314, 1979.

[Schumer 1980] Schumer, Robert A., "Mechanisms in human stereopsis," Ph.D. thesis, Department of Psychology, Stanford University, 1979.

[Shafer 1980] Shafer, Steven A., "MOOSE. Users' Manual, Implementation Guide, Evaluation," Bericht 70, Report IfI-HH-B-70/80, Fachbereich Informatik, Universität Hamburg, April 1980.

[Shanmugam 1979] Shanmugam, K.S., F.M. Dickey, J.A. Green, "An optimal frequency domain filter for edge detection in digital images," *IEEE Trans Pattern Analysis and Machine Intelligence*, PAMI-1, Jan. 1979, 39–47.

[Shapiro 1974] Shapiro, S.D., "Detection of lines in noisy pictures using clustering," *Proceedings of the Second International Joint Conference on Pattern Recognition*, Copenhagen, Aug. 13–15, 1974, 317–318.

[Shapiro 1975] Shapiro, S.D., "Transformations for the Computer Detection of Curves in Noisy Pictures," *Computer Graphics and Image Processing*, 4, 1975, p. 328.

[Shapiro 1978] Shapiro, S.D., "Generalization of the Hough Transform for Curve Detection in Noisy Digital Images," *Proceedings of the Fourth International Joint Conference on Pattern Recognition* (IJCPR-78), 710–714.

[Shaw 1977] Shaw, G.B.,"Local and Regional Edge Detectors: Some Comparisons," Univ. of Maryland Technical Report TR-614, December 1977.

[Shaw 1979] Shaw, G.B.,"Local and Regional Edge Detectors: Some Comparisons," Computer Graphics and Image Processing, vol. 9, no. 2, Feb. 1979, 135–149.

[Shirai 1975] Shirai, Y., "Edge finding, segmentation of edges and recognition of complex objects," Proc. 4th IJCAI, 1975, 674-681.

[Shirai 1978] Shirai, Y., "Recognition of man-made objects using edge cues," **Computer Vision Systems**, A. Hanson, E. Riseman, eds, Academic Press, New York, 1978.

[Slama 1980] Slama, C.C., editor-in-chief, *Manual of Photogrammetry*, American Society of Photogrammetry, 1980.

[Slepian 1961] Slepian, D. and H.O. Pollak, "Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty — I," *Bell Syst. Tech. J.*, 40, January 1961, 43-63.

[Smith 1977] Smith, Babington, "A wartime anticipation of random-dot stereograms," *Perception*, Vol. 6(2), 233-234, 1977.

[Sobel 1970] Sobel, Irwin, "Camera models and machine perception," Stanford University Report, AIM-121, 1970.

[Somerville 1976] Somerville, C. and J.L. Mundy, "One Pass Contouring of Images Through Planar Approximation," *Proc. of the 3rd International Joint Conference on Pattern Recognition (IJCPR-76)*, Nov. 1976 (IEEE 76CH1140-3C).

[Sugie 1976] Sugie, N. and Suwa, M., "Notes on vision research. V. A model of binocular depth perception suggested by neurophysiological evidence," *Bull. Electrotech. Lab. (Japan)*, Vol. 40(11), 890-921, 1976.

[Tamura 1977] Tamura, H., S. Mori, and T. Yamawaki, "Psychological and Computational Measurements of Basic Textural Features and Their Comparison," ETL and Waseda University, Japan, 1977.

[Tamura 1978] Tamura, H., S. Mori and T. Yamawaki, "Textural Features Corresponding to Visual Perception, " *IEEE Transactions on Systems, Man and Cybernetics*, vol. 8, no. 6, June 1978, 460–473.

[Thacker 1979] Thacker, McCreight, Lampson, Sproull, Boggs, "Alto: A Personal Computer," CSL-79-11, Xerox PARC, August 1979.

[Theophrastus 400BC] Theophrastus, "De Sensu," 61-2.

[Thomas 1974] Thomas, A.J., and T.O. Binford, "Information processing Analysis of Visual Perception: A Review," Stanford Artificial Intelligence Laboratory, AIM-227, June 1974.

[Thompson 1977] Thompson, W., "Textural Boundary Analysis," *IEEE Transactions on Computers*, vol. 26, 1977, 272–276.

[Thompson 1981] Thompson, William B. "Combining Motion and Contrast for Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, November 1980.

[Tomita 1979] Tomita, F., Y. Shirai and S. Tsuji, "Description of Textures by a Structural Analyzer," *Proceedings of the International Joint Conference on Artificial Intelligence*, Tokyo, August 1979, 884–889.

[Tsuji 1981] Saburo Tsuji, Michiharu Osada, and Masahiko Yachida, "Tracking and Segmentation of Moving Objects in Dynamic Line Images," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 2, no. 6, Nov 1980.

[Turner 1974] Turner, K., "Computer Perception of curved objects using a television camera," Ph.D. dissertation, Edinburgh University, November 1974.

[Tyler 1979] Tyler, Christopher and Sutter, Erich, "Depth from spatial frequency difference: an old kind of stereopsis?," *Vision Research*, Vol. 19(8), 859-865, 1979.

[Uttal 1975] Uttal, William, Fitzgerald, Judy, Eskin, Thelma, "Rotation and translation effects on stereoscopic acuity," *Vision Research*, Vol. 15, (8-9) 939-944, 1975.

[Victor 1978] Victor, J.D. and S. Brodie, "Discriminable Textures with Identical Buffon Needle Statistics," *Biological Cybernetics*, vol. 31,no. 4, 1978, 231–234.

[Vilnrotter 1980] Vilnrotter, F., R. Nevatia and K. Price, "Structural Description of Natural Textures," *Proceedings of Fifth International Pattern Recognition Conference*, Miami, Dec. 1980.

[Vilnrotter 1981] Vilnrotter, F., "Structural Analysis of Natural Textures," University of Southern California, Ph. D. Thesis, USCISG 100, September 1981.

[Walk 1961] Walk, Richard D. and Eleanor J. Gibson, "A Comparative and Analytic Study of Visual Depth Perception," *Psychological Monographs*, vol. 75, no. 15, 2–34, 1961.

[Wallach 1976] Wallach, Hans and Bacon, Joshua, "Two forms of retinal disparity," *Perception and Psychophysics*, vol, 19(5) 375-382, 1976.

[Waller 1981] Waller, Larry, "LISP Language Gets Special Machine," *Electronics*, Vol. 54, No. 17, August 1981.

[Watson 1982] Watson, Ian, and John Gurd, "A Practical Data Flow Computer," *Computer*, Vol. 15, No. 2, February 1982.

[Webb 1981] Webb, J. and J.K. Agarwal, "Structure from motion of rigid and jointed objects," *Seventh Int. Joint Conf. on Artificial Intelligence*, Vancouver, B.C., 686–691, August 1981.

[Weszka 1976] Weszka, J., C.R. Dyer and A. Rosenfeld, "A Comparative Study of Texture Measures for Terrain Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6, no. 4, April 1976, 269–285.

[Will 1971] Will, P.M., and K.S. Pennington, "Grid Coding: A Novel Technique for Image Processing," IBM RC 3456, July 1971.

[Williams 1981] Williams, Thomas D., "Depth from Camera Motion in a Real World Scene," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 6, November 1980.

[Wilson 1977] Wilson, H.R., Giese, S.C., "Threshold visibility of frequency gradient patterns," *Vision Res.* 17, 1177-1190, 1977.

[Wilson 1978] Wilson, Hugh R., James R. Bergen, "A four mecahanism model for threshold spatial vision," University of Chicago, April 1978.

[Wilson 1978] Wilson, Hugh R., "Quantitative Prediction of Line Spread Function Measurements: Implications for Channel Bandwidths," *Vision Research*, vol. 18, 493–496, 1978.

[Wist 1971] Wist, E.R. and Freund, H.J., "The neuronal basis of binocular vision," *Pattern Recognition in Biological and Technical Systems*, 288-300, 1971.

[Wong 1975] Wong, K.W., "Mathematical Formulation and digital analysis in close-range photometry," *Photogrammetric Engineering and Remote Sensing*, vol. 41, 1355, 1975.

[Yakimovsky 1976] Yakimovsky, Y., "Boundary and Object Detection in Real World Images," *Journal of the ACM*, vol. 23, no. 4, October 1976, p. 599.

[Yellott 1979] Yellott, John and Kaiwi, Jerry, "Depth inversion despite stereopsis: the appearance of random-dot stereograms on surfaces seen in reverse perspective," *Perception*, Vol. 8(2), 135-142, 1979.

[Yonas 1978] Yonas, A., Wallace T. Cleaves, and Linda Pettersen, "Development of Sensitivity to Pictorial Depth," *Science*, Vol. 200, 77-79, April 1978.

[Young 1978] Young, W.H. and D.M. Isabell, "Production Mapping with Orthophoto Digital Terrain Models," *Photogrammetric Engineering and Remote Sensing*, vol. 44, no. 12, 1521–1536, December 1978.

[Zobrist 1975] Zobrist, A. and W. Thompson, "Building a Distance Function for Gestalt Grouping," *IEEE Transactions on Computers*, vol. 24, 1975, 718–728.

[Zucker 1976] Zucker, S.W., "Toward a Model of Texture," *Computer Graphics and Image Processing*, vol. 5, 1976, 190–202.

[Zucker 1977] Zucker, S.W., R.A. Hummel, and A. Rosenfeld, "An application of relaxation labelling to line and curve enhancement," *IEEE Trans. Computers* C-26, 1977, 394–403.

addresses	number of copies
John T. Boland RADC/IRRA	20
RADC/TSLD GRIFFISS AFB NY 13441	1
RADC/DAP GRIFFISS AFB NY 13441	2
ADMINISTRATOR DEF TECH INF CTR ATTN: DTIC-DDA CAMERON STA BG 5 ALEXANDRIA VA 22314	12
HQ ESC (XPZP) SAN ANTONIO TX 78243	1
HQ ESC/DOO SAN ANTONIO TX 78243	1
DMAAC/STT ST LOUIS AFS MO 63125	1
DMAHTC/STT 6500 Brooks Lane WASHINGTON DC 20315	2

HQ USAF/XOKT WASHINGTON DC 20330	1
HQ USAF/RDST WASHINGTON DC 20330	1
HQ USAF/RDPV WASHINGTON DC 20330	1
DIRECTOR DMAHTC ATTN: SDSIM 6500 Brookes Lane WASH DC 20315	1
RADC/RBRAC GRIFFISS AFB NY 13441	1
PENTAGON USDR&E, RM 3D-139 ATTN: TSCO WASHINGTON DC 20301	2
HQ AFSC/DLAE ANDREWS AFB DC 20334	1
HQ AFSC/SDWI ANDREWS AFB DC 20334	1
HQ AFSC/SDWR ANDREWS AFB DC 20334	1

HQ AFSC/XRPA ANDREWS AFB DC 20334	1
HQ AFSC/XRK ANDREWS AFB DC 20334	1
HQ SAC/NRI (STINFO LIBRARY) OFFUTT AFB NE 68113	1
HQ 3246 TW/TETE EGLIN AFB FL 32542	1
HQ 3246 TW/TETW EGLIN AFB FL 32542	1
AFATL/DLODL EGLIN AFB FL 32542	1
TAFIG/IIPE LANGLEY AFB VA 23665	1
HQ TAC/XPS (STINFO)] LANGLEY AFB VA 23665	1
HQ TAC/XPJC LANGLEY AFB VA 23665	1
TAFIG/IICJ ATTN: Capt John Morrison LANGLEY AFB VA 23665	2

HQ TAC/DRCC LANGLEY AFB VA 23665	1
HQ TAC/DRF LANGLEY AFB VA 23665	1
AFSC LIAISON OFFICE LANGLEY RESEARCH CENTER (NASA) LANGLEY AFB VA 23665	1
AFWL/NTYEE (C E BAUM) KIRTLAND AFB NM 87117	1
AFWL/SUL ATTN: TECHNICAL LIBRARY KIRTLAND AFB NM 87117	1
ASD/ENEGE ATTN: CAPT T CLELAND WRIGHT-PATTERSON AFB OH 45433	1
ASD/ENEGE ATTN: MR LARRY WEAVER WRIGHT-PATTERSON AFB OH 45433	1
ASD/AEI (PLAISTED) WRIGHT-PATTERSON AFB OH 45433	1
ASD/XRM (MR R H SUDHEIMER/F. RATH) WRIGHT-PATTERSON AFB OH 45433	1

ASD/XRS 1
WRIGHT-PATTERSON AFB OH 45433

1

AFIT/LDE - TECHNICAL LIBRARY 1
BUILDING 640, AREA B
WRIGHT-PATTERSON AFB OH 45433

AFHRL/LRS 1
WRIGHT-PATTERSON AFB OH 45433

ASD-AFALD/AXT 1
WRIGHT-PATTERSON AFB OH 45433

AFHRL/OTN 1
WILLIAMS AFB AZ 85224

AFHRL/OTS 1
Williams AFB AZ 85224

AUL/LSE 67-342 1
MAXWELL AFB AL 36112

HQ AFCC/DAPL 1
BLDG P-40 NORTH, RM 9
SCOTT AFB IL 62225

AFHRL/LRT 1
LOWRY AFB CO 80230

3420 TCHTG/TTMNL 1
LOWRY AFB CO 80230

3300 TTW/TTGX 1
KEESLER AFB MS 39534

DEFENSE INTELLIGENCE AGENCY 1
ATTN: RSE-2 (LT COL SCHWARTZ)
WASHINGTON DC 20301

CODE R123 TECHNICAL LIBRARY 1
DEFENSE COMMUNICATIONS
ENGINEERING CENTER
1860 WIEHLE AVENUE
RESTON VA 22090

DIRECTOR 1
DEFENSE NUCLEAR AGENCY
ATTN: TIL
WASHINGTON DC 20305

CHIEF, C3 DIVISION 2
DEVELOPMENT CENTER, MCDEC
ATTN: R S HARTMAN
QUANTICA VA 22134

AFLMC/LGY 1
ATTN: MAJOR MORGAN
GUNTER AFS AL 36114

DIRECTOR 1
BMD ADVANCED TECHNOLOGY CENTER
ATTN: ATC-P, CHARLES VICK
PO BOX 1500
HUNTSVILLE AL 35807

EOARD/CMI 1
TECHNICAL LIBRARY FL 2878
BOX 14
FPO NY 09510

COMMANDING OFFICER 1
NAVAL AVIONICS CENTER
LIBRARY - CODE 765
INDIANAPOLIS IN 46218

NAVAL TRAINING EQUIPMENT CENTER 1
TECHNICAL INFORMATION CENTER
ORLANDO FL 32813

COMMANDER 1
NAVAL OCEAN SYSTEMS CENTER
ATTN: TECHNICAL LIBRARY, CODE 4473B
SAN DIEGO CA 92152

SUPERINTENDENT (CODE 1424) 1
NAVAL POSTGRADUATE SCHOOL
MONTEREY CA 93940

COMMANDING OFFICER 1
NAVAL RESEARCH LABORATORY
CODE 2627
WASHINGTON DC 20375

REDSTONE SCIENTIFIC INFORMATION CENTER 2
ATTN: DRSMI-RPRD
US ARMY MISSILE COMMAND
REDSTONE ARSENAL AL 35809

MILITARY SEALIFT COMMAND 1
TECHNICAL INFORMATION CENTER< M-00T6
DEPARTMENT OF THE NAVY
WASH DC 20390

DOT/FAA TECHNICAL CENTER 1
ARD-142 (ATTN: A R CIOFFI)
ATLANTIC CITY NJ 08405

NATIONAL CENTER FOR ATMOSPHERIC RESEARCH 1
MESA LIBRARY
PO BOX 3000
BOULDER CO 80307

FRANK J SEILER RESEARCH LAB 1
FJSRL/NHL
US AIR FORCE ACADEMY CO 80840

AIR FORCE ELEMENT (AFELM) 1
THE RAND CORP
1700 MAIN STREET
SANTA MONICA CA 90406

DR RAYNER K ROSICH 1
ELECTRO MAGNETIC APPLNS, INC
C/O 7031 PIERSON STREET
ARVADE CO 80004

AEDC LIBRARY (TECH FILES) 1
ARNOLD AFS TN 37389

Director 1
National Security Agency
ATTN: W07
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: W22
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: W31
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: S809 (Mr. Haenchen)
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: R1
Fort Meade MD 20755

Director 1
National Security
ATTN: R2
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: R5
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: R091
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: R7
Fort Meade MD 20755

Director 1
National Security Agency
ATTN: R8
Fort Meade MD 20755

HQ ESD/FAE, STOP 27 1
HANSCOM AFB MA 01731

ESD/TCBR 1
HANSCOM AFB MA 01731

ESD/DCKD (STOP 53) 1
ATTN: LT COMBS
HANSCOM AFB MA 01731

ESD/XRT 1
HANSCOM AFB MA 01731

ESD/XRV HANSCOM AFB MA 01731	1
ESD/XRC HANSCOM AFB MA 01731	1
ESD/XRG HANSCOM AFB MA 01731	1
ESD/XRTR HANSCOM AFB MA 01731	1
ESD/TCG HANSCOM AFB MA 01731	1
ESD/XR HANSCOM AFB MA 01731	2
ESD/DCR-3E HANSCOM AFB MA 01731	1
HQ ESD/YSM (STOP 18) HANSCOM AFB MA 01731	2
HQ ESD/DCR-1S Hanscom AFB MA 01731	1

HQ ESD/DCR-1I 1
HANSCOM AFB MA 01731

DEPARTMENT OF TRANSPORTATION 1
LIBRARY, 10A SVS BR, M494.6
800 INDEPENDENCE AVE, S.W.
WASH DC 20591

AFEWC/ESRI 1
San Antonio TX 78243

485 EIG/EIEXR (DMO) 2
Griffiss AFB NY 13441

Stanford University AI Laboratory 5
ATTn. Thomas O. Binford
COmputer Science Department
Stanford University
Stanford CA 94305

John Entzminger, Tecnical Director 1
Intelligence and Reconnaissance Division
Rome Air Development Center
Griffiss AFB, NY 13441

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.